COFFEE QUALITY
1 Milestone-1 Evaluation Project

Project Documentation: Exploratory Data Analysis of Coffee quality
Dataset :
Title : Data Analysis on Coffee quality
Name : Praveen kumar N
DA/DS : Data Analytics (DA)
Batch number : B4 (June - Online)(M) - DA & DS
Online/Offline : Online
Roll Number : 60624OLR013

Table of Contents :

NOTE : All the codes used for this are given after the documentation and displaying of results.

1) Introduction :

The coffee quality dataset comprises various attributes related to coffee beans, including Species , flavor , aroma , etc. The goal of this project is to conduct a comprehensive analysis of the dataset to derive insights into coffee quality as to what actually derives the quality of a coffee which can be helpful for our clients whether they are thinking of starting a new coffee brand or even for normal consumers to choose the best brand of coffee to have the best cup of coffee.

Columns in the dataset related to Coffee quality:

- Species:Species of the coffee plant , here, only arabica and robusta is present.
- Owner:The one who farmed the coffee plants.
- Country.of.Origin:From which country it comes
- Farm.Name: The farm the coffee was grown
- Lot.Number: The number of the lot
- Mill: The mill it was grown
- ICO.Number: It is unique identifier asssigned to each bag of coffee bean
- Company: The company that imports and export the coffee beans.
- Altitude: How much above the sea level the coffee farm is situated
- Region: The region where the coffee plants were grown.
- Producer: The one who produces which all coffee.
- Number.of.Bags: Number of bags produced.
- Bag.Weight: Weight of 1 bag.
- In.Country.Partner: The country in which the supplier is for specific coffee company.
- Harvest.Year: The year it was harvested.
- Grading.Date: The date it was graded.O
- Owner.1: The person who got it right after first exporting.
- Variety: Variety of the coffee bean

- Processing.Method: The method coffee bean was processed(like washed , semi-washed , natural ,etc)
- Aroma: How good the smell is.
- Flavor: How good the flavor of the coffee is.
- Aftertaste: The aftertaste that the coffee leaves in your mouth
- Acidity: How low the pH of the coffee is.
- Body: Refers to the texture and weight of the coffee in your mouth.
- Balance: Refers to harmony and equilibrium of flavors,acidity and body in a cup of coffee.
- Uniformity: Refers to the consistency of flavor, quality, and appearance of the coffee beans
- Clean.Cup: Refers to a cup of coffee that is free from defects, impurities, and off-flavors
- Sweetness: How sweet the coffee is
- Cupper.Points: Refer to a standardized system used to evaluate and score the quality of coffee
- Total.Cup.Points: Refers to the final score assigned to a coffee based on the evaluation of its various attributes and using the cupper point
- Moisture: Amount of moisture the coffee has retained.
- Category.One.Defects: More severe defects that affect coffee quality(like, Moldy, Skunky, fermented, etc)
- Quakers: Refer to a type of defective coffee bean that is lighter in color and has a distinct flavor.
- Color: Refers to the visual appearance of the coffee beans which can indicate various aspects of coffee.
- Category.Two.Defects: Less severe defects that affect coffee quality(like, woody, nutty, papery, etc)
- Expiration: When the coffee becomes unable to consume.

- Certification.Body: Organisation that ensure that coffee beans, farms, or production processes meet certain standards.
- Certification.Address: Address of the certification body that certified the specific bag of coffee beans.
- Certification.Contact: Contact method and info for the certification body.
- unit_of_measurement: Units that help coffee professionals and enthusiasts measure, communicate, and perfect their coffee-related tasks.
- altitude_low_meters: Refers to a measurement of altitude (height above sea level) that is relatively low, in meters.
- altitude_high_meters: Refers to a measurement of altitude (height above sea level) that is relatively high, in meters.
- altitude_mean_meters: Altitude mean meters refers to the average height of a location or area above sea level, measured in meters.

2) Aim :

The aim of this project is to conduct a comprehensive analysis of the dataset to derive insights into overall coffee quality, catering to both consumers and manufacturers in the computer industry.

3)Problem Statement:

The coffee market is highly competitive, as almost everyone consumes coffee, most of us need a cup of coffee to even function properly.So it is essential for someone trying to make their own coffee brand to aware what all to focus on to make sure that they have the best quality of coffee in the current market and what all to keep an eye on to stay ahead of the competition. This can also be useful for coffee enthusiasts to make sure that they start their day with the best quality of coffee.

4) Project Workflow :

Overview of the project workflow or methodology followed.

- Data Cleaning
- Exploratory Data Analysis (EDA)
- Data Visualization
- Analysis and Interpretation
- Documentation

5) Data Understanding :

Description of the dataset, including structure, dimensions, and data types.

- There are 1339 rows and 44 columns in the Dataset.
- From the info we conclude that out of the 44 columns, 24 were object type, 17 were float and 3 were integer.
- Unnamed: 0 column should be dropped

```
import numpy as np
import pandas as pd
df=pd.read_csv("D:/KGISL MICRO COLL/MIlestone 1/coffeeQuality.csv")
#Loading data from a CSV file into a Pandas DataFrame
df
```

```
      Unnamed: 0  Species                    Owner Country_of_Origin
\
0              0  Arabica                metad plc          Ethiopia

1              1  Arabica                metad plc          Ethiopia

2              2  Arabica  grounds for health admin         Guatemala

3              3  Arabica       yidnekachew dabessa          Ethiopia

4              4  Arabica                metad plc          Ethiopia

...          ...      ...                      ...               ...

1334        1334  Robusta               luis robles           Ecuador

1335        1335  Robusta               luis robles           Ecuador

1336        1336  Robusta               james moore     United States

1337        1337  Robusta             cafe politico             India

1338        1338  Robusta             cafe politico           Vietnam


                                     Farm.Name Lot.Number
Mill  \
0                                    metad plc        NaN       metad
plc
1                                    metad plc        NaN       metad
plc
2     san marcos barrancas "san cristobal cuch        NaN
NaN
3        yidnekachew dabessa coffee plantation        NaN
wolensu
4                                    metad plc        NaN       metad
plc
...                                        ...        ...
...
1334                                 robustasa   Lavado 1     our own
lab
1335                                 robustasa   Lavado 3  own
laboratory
1336                          fazenda cazengo        NaN    cafe
cazengo
```

```
1337                                        NaN        NaN
NaN
1338                                        NaN        NaN
NaN

             ICO.Number                                  Company
Altitude  \
0             2014/2015      metad agricultural developmet plc
1950-2200
1             2014/2015      metad agricultural developmet plc
1950-2200
2                   NaN                                   NaN  1600 -
1800 m
3                   NaN  yidnekachew debessa coffee plantation
1800-2200
4             2014/2015      metad agricultural developmet plc
1950-2200
...                 ...                                   ...
...
1334                NaN                             robustasa
NaN
1335                NaN                             robustasa
40
1336                NaN                 global opportunity fund     795
meters
1337  14-1118-2014-0087                          cafe politico
NaN
1338                NaN                          cafe politico
NaN

      ...       Color Category.Two.Defects           Expiration  \
0     ...       Green                    0      April 3rd, 2016
1     ...       Green                    1      April 3rd, 2016
2     ...         NaN                    0       May 31st, 2011
3     ...       Green                    2     March 25th, 2016
4     ...       Green                    2      April 3rd, 2016
...   ...         ...                  ...                  ...
1334  ...  Blue-Green                    1   January 18th, 2017
1335  ...  Blue-Green                    0   January 18th, 2017
1336  ...         NaN                    6  December 23rd, 2015
1337  ...       Green                    1    August 25th, 2015
1338  ...         NaN                    9    August 25th, 2015

                     Certification.Body  \
0     METAD Agricultural Development plc
1     METAD Agricultural Development plc
2           Specialty Coffee Association
3     METAD Agricultural Development plc
4     METAD Agricultural Development plc
...                                  ...
```

```
1334        Specialty Coffee Association
1335        Specialty Coffee Association
1336        Specialty Coffee Association
1337        Specialty Coffee Association
1338        Specialty Coffee Association

                         Certification.Address  \
0     309fcf77415a3661ae83e027f7e5f05dad786e44
1     309fcf77415a3661ae83e027f7e5f05dad786e44
2     36d0d00a3724338ba7937c52a378d085f2172daa
3     309fcf77415a3661ae83e027f7e5f05dad786e44
4     309fcf77415a3661ae83e027f7e5f05dad786e44
...                                        ...
1334  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1335  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1336  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1337  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1338  ff7c18ad303d4b603ac3f8cff7e611ffc735e720

                         Certification.Contact unit_of_measurement  \
0     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
1     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
2     0878a7d4b9d35ddbf0fe2ce69a2062cceb45a660                    m
3     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
4     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
...                                        ...                  ...
1334  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1335  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1336  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1337  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1338  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m

     altitude_low_meters altitude_high_meters altitude_mean_meters
0                 1950.0               2200.0                2075.0
1                 1950.0               2200.0                2075.0
2                 1600.0               1800.0                1700.0
3                 1800.0               2200.0                2000.0
4                 1950.0               2200.0                2075.0
...                  ...                  ...                   ...
1334                 NaN                  NaN                   NaN
1335                40.0                 40.0                  40.0
1336               795.0                795.0                 795.0
1337                 NaN                  NaN                   NaN
1338                 NaN                  NaN                   NaN

[1339 rows x 44 columns]

df.head()
```

```
   Unnamed: 0  Species                    Owner Country_of_Origin  \
0           0  Arabica                metad plc          Ethiopia
1           1  Arabica                metad plc          Ethiopia
2           2  Arabica  grounds for health admin         Guatemala
3           3  Arabica       yidnekachew dabessa          Ethiopia
4           4  Arabica                metad plc          Ethiopia

                                  Farm.Name Lot.Number       Mill
ICO.Number  \
0                                 metad plc        NaN  metad plc
2014/2015
1                                 metad plc        NaN  metad plc
2014/2015
2  san marcos barrancas "san cristobal cuch        NaN        NaN
NaN
3     yidnekachew dabessa coffee plantation        NaN    wolensu
NaN
4                                 metad plc        NaN  metad plc
2014/2015

                                   Company       Altitude  ...  Color  \
0      metad agricultural developmet plc      1950-2200  ...  Green
1      metad agricultural developmet plc      1950-2200  ...  Green
2                                    NaN  1600 - 1800 m  ...    NaN
3  yidnekachew debessa coffee plantation      1800-2200  ...  Green
4      metad agricultural developmet plc      1950-2200  ...  Green

  Category.Two.Defects        Expiration
Certification.Body  \
0                    0   April 3rd, 2016  METAD Agricultural
Development plc
1                    1   April 3rd, 2016  METAD Agricultural
Development plc
2                    0    May 31st, 2011        Specialty Coffee
Association
3                    2  March 25th, 2016  METAD Agricultural
Development plc
4                    2   April 3rd, 2016  METAD Agricultural
Development plc

                      Certification.Address  \
0  309fcf77415a3661ae83e027f7e5f05dad786e44
1  309fcf77415a3661ae83e027f7e5f05dad786e44
2  36d0d00a3724338ba7937c52a378d085f2172daa
3  309fcf77415a3661ae83e027f7e5f05dad786e44
4  309fcf77415a3661ae83e027f7e5f05dad786e44

                      Certification.Contact unit_of_measurement  \
0  19fef5a731de2db57d16da10287413f5f99bc2dd                    m
1  19fef5a731de2db57d16da10287413f5f99bc2dd                    m
```

```
2  0878a7d4b9d35ddbf0fe2ce69a2062cceb45a660                    m
3  19fef5a731de2db57d16da10287413f5f99bc2dd                    m
4  19fef5a731de2db57d16da10287413f5f99bc2dd                    m

  altitude_low_meters altitude_high_meters altitude_mean_meters
0              1950.0               2200.0               2075.0
1              1950.0               2200.0               2075.0
2              1600.0               1800.0               1700.0
3              1800.0               2200.0               2000.0
4              1950.0               2200.0               2075.0

[5 rows x 44 columns]
```

```
df.tail()
```

```
      Unnamed: 0  Species          Owner Country_of_Origin
Farm.Name  \
1334        1334  Robusta    luis robles           Ecuador
robustasa
1335        1335  Robusta    luis robles           Ecuador
robustasa
1336        1336  Robusta    james moore     United States  fazenda
cazengo
1337        1337  Robusta  cafe politico             India
NaN
1338        1338  Robusta  cafe politico           Vietnam
NaN

     Lot.Number            Mill         ICO.Number
Company  \
1334   Lavado 1     our own lab                NaN
robustasa
1335   Lavado 3  own laboratory                NaN
robustasa
1336        NaN    cafe cazengo                NaN  global opportunity
fund
1337        NaN             NaN  14-1118-2014-0087            cafe
politico
1338        NaN             NaN                NaN            cafe
politico

       Altitude  ...       Color Category.Two.Defects
Expiration  \
1334         NaN  ...  Blue-Green                   1   January 18th,
2017
1335          40  ...  Blue-Green                   0   January 18th,
2017
1336  795 meters  ...         NaN                   6  December 23rd,
2015
1337         NaN  ...       Green                   1    August 25th,
```

```
2015
1338         NaN  ...         NaN                    9    August 25th,
2015

                Certification.Body
Certification.Address  \
1334  Specialty Coffee Association
ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1335  Specialty Coffee Association
ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1336  Specialty Coffee Association
ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1337  Specialty Coffee Association
ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1338  Specialty Coffee Association
ff7c18ad303d4b603ac3f8cff7e611ffc735e720

                        Certification.Contact unit_of_measurement  \
1334  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1335  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1336  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1337  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1338  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m

     altitude_low_meters altitude_high_meters altitude_mean_meters
1334                 NaN                  NaN                  NaN
1335                40.0                 40.0                 40.0
1336               795.0                795.0                795.0
1337                 NaN                  NaN                  NaN
1338                 NaN                  NaN                  NaN

[5 rows x 44 columns]

df.describe()

        Unnamed: 0  Number_of_Bags        Aroma       Flavor
Aftertaste  \
count  1339.000000     1338.000000  1339.000000  1339.000000
1339.000000
mean    669.000000      159.085202     7.770187     7.520426
7.401083
std     386.680316      173.698167     5.534440     0.398442
0.404463
min       0.000000        0.000000     0.000000     0.000000
0.000000
25%     334.500000       14.000000     7.420000     7.330000
7.250000
50%     669.000000      175.000000     7.580000     7.580000
7.420000
75%    1003.500000      275.000000     7.750000     7.750000
```

```
7.580000
max    1338.000000     3200.000000   200.000000     8.830000
8.670000

           Acidity         Body      Balance   Uniformity    Clean.Cup
\
count  1339.000000  1339.000000  1339.000000  1339.000000  1339.000000

mean      7.535706     7.517498     7.518013     9.834877     9.835108

std       0.379827     0.370064     0.408943     0.554591     0.763946

min       0.000000     0.000000     0.000000     0.000000     0.000000

25%       7.330000     7.330000     7.330000    10.000000    10.000000

50%       7.580000     7.500000     7.500000    10.000000    10.000000

75%       7.750000     7.670000     7.750000    10.000000    10.000000

max       8.750000     8.580000     8.750000    10.000000    10.000000


         Sweetness  Cupper.Points  Total.Cup.Points     Moisture  \
count  1339.000000    1339.000000       1339.000000  1339.000000
mean      9.856692       7.503376         82.089851     0.088379
std       0.616102       0.473464          3.500575     0.048287
min       0.000000       0.000000          0.000000     0.000000
25%      10.000000       7.250000         81.080000     0.090000
50%      10.000000       7.500000         82.500000     0.110000
75%      10.000000       7.750000         83.670000     0.120000
max      10.000000      10.000000         90.580000     0.280000

       Category.One.Defects      Quakers  Category.Two.Defects  \
count           1339.000000  1338.000000           1339.000000
mean               0.479462     0.173393              3.556385
std                2.549683     0.832121              5.312541
min                0.000000     0.000000              0.000000
25%                0.000000     0.000000              0.000000
50%                0.000000     0.000000              2.000000
75%                0.000000     0.000000              4.000000
max               63.000000    11.000000             55.000000

       altitude_low_meters  altitude_high_meters  altitude_mean_meters

count          1109.000000           1109.000000           1109.000000

mean           1750.713315           1799.347775           1775.030545

std            8669.440545           8668.805771           8668.626080
```

```
min                1.000000              1.000000              1.000000

25%             1100.000000           1100.000000           1100.000000

50%             1310.640000           1350.000000           1310.640000

75%             1600.000000           1650.000000           1600.000000

max           190164.000000         190164.000000         190164.000000
```

```
#Displays a concise summary of the DataFrame's structure, content, and memory usage
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1339 entries, 0 to 1338
Data columns (total 44 columns):
 #   Column                 Non-Null Count  Dtype  
---  ------                 --------------  -----  
 0   Unnamed: 0             1339 non-null   int64  
 1   Species                1339 non-null   object 
 2   Owner                  1332 non-null   object 
 3   Country_of_Origin      1338 non-null   object 
 4   Farm.Name              980 non-null    object 
 5   Lot.Number             276 non-null    object 
 6   Mill                   1021 non-null   object 
 7   ICO.Number             1180 non-null   object 
 8   Company                1130 non-null   object 
 9   Altitude               1113 non-null   object 
 10  Region                 1280 non-null   object 
 11  Producer               1107 non-null   object 
 12  Number_of_Bags         1338 non-null   float64
 13  Bag.Weight             1339 non-null   object 
 14  In.Country.Partner     1339 non-null   object 
 15  Harvest.Year           1292 non-null   object 
 16  Grading.Date           1339 non-null   object 
 17  Owner_1                1332 non-null   object 
 18  Variety                1113 non-null   object 
 19  ProcessingMethod       1169 non-null   object 
 20  Aroma                  1339 non-null   float64
 21  Flavor                 1339 non-null   float64
 22  Aftertaste             1339 non-null   float64
 23  Acidity                1339 non-null   float64
 24  Body                   1339 non-null   float64
 25  Balance                1339 non-null   float64
 26  Uniformity             1339 non-null   float64
 27  Clean.Cup              1339 non-null   float64
 28  Sweetness              1339 non-null   float64
 29  Cupper.Points          1339 non-null   float64
```

```
 30  Total.Cup.Points       1339 non-null   float64
 31  Moisture               1339 non-null   float64
 32  Category.One.Defects   1339 non-null   int64
 33  Quakers                1338 non-null   float64
 34  Color                  1069 non-null   object
 35  Category.Two.Defects   1339 non-null   int64
 36  Expiration             1339 non-null   object
 37  Certification.Body     1339 non-null   object
 38  Certification.Address  1339 non-null   object
 39  Certification.Contact  1339 non-null   object
 40  unit_of_measurement    1339 non-null   object
 41  altitude_low_meters    1109 non-null   float64
 42  altitude_high_meters   1109 non-null   float64
 43  altitude_mean_meters   1109 non-null   float64
dtypes: float64(17), int64(3), object(24)
memory usage: 460.4+ KB
```

```
df.columns
```

```
Index(['Unnamed: 0', 'Species', 'Owner', 'Country_of_Origin',
'Farm.Name',
       'Lot.Number', 'Mill', 'ICO.Number', 'Company', 'Altitude',
'Region',
       'Producer', 'Number_of_Bags', 'Bag.Weight',
'In.Country.Partner',
       'Harvest.Year', 'Grading.Date', 'Owner_1', 'Variety',
       'ProcessingMethod', 'Aroma', 'Flavor', 'Aftertaste', 'Acidity',
'Body',
       'Balance', 'Uniformity', 'Clean.Cup', 'Sweetness',
'Cupper.Points',
       'Total.Cup.Points', 'Moisture', 'Category.One.Defects',
'Quakers',
       'Color', 'Category.Two.Defects', 'Expiration',
'Certification.Body',
       'Certification.Address', 'Certification.Contact',
'unit_of_measurement',
       'altitude_low_meters', 'altitude_high_meters',
'altitude_mean_meters'],
      dtype='object')
```

```
#Removing the unnamed column
a= df.drop(df.columns[[0]], axis=1)
```

```
a
```

```
      Species                     Owner Country_of_Origin  \
0     Arabica                 metad plc          Ethiopia
1     Arabica                 metad plc          Ethiopia
2     Arabica  grounds for health admin         Guatemala
3     Arabica       yidnekachew dabessa          Ethiopia
```

```
4     Arabica                  metad plc          Ethiopia
...       ...                        ...               ...
1334  Robusta                luis robles           Ecuador
1335  Robusta                luis robles           Ecuador
1336  Robusta                james moore     United States
1337  Robusta              cafe politico             India
1338  Robusta              cafe politico           Vietnam

                                     Farm.Name Lot.Number
Mill  \
0                                    metad plc        NaN       metad
plc
1                                    metad plc        NaN       metad
plc
2     san marcos barrancas "san cristobal cuch        NaN
NaN
3        yidnekachew dabessa coffee plantation        NaN
wolensu
4                                    metad plc        NaN       metad
plc
...                                        ...        ...
...
1334                                 robustasa   Lavado 1     our own
lab
1335                                 robustasa   Lavado 3  own
laboratory
1336                           fazenda cazengo        NaN    cafe
cazengo
1337                                       NaN        NaN
NaN
1338                                       NaN        NaN
NaN

             ICO.Number                                  Company
Altitude  \
0             2014/2015      metad agricultural developmet plc
1950-2200
1             2014/2015      metad agricultural developmet plc
1950-2200
2                   NaN                                   NaN  1600 -
1800 m
3                   NaN  yidnekachew debessa coffee plantation
1800-2200
4             2014/2015      metad agricultural developmet plc
1950-2200
...                 ...                                   ...
...
1334                NaN                              robustasa
NaN
```

```
1335                NaN                                robustasa
40
1336                NaN                global opportunity fund     795
meters
1337  14-1118-2014-0087                          cafe politico
NaN
1338                NaN                          cafe politico
NaN

                             Region  ...       Color
Category.Two.Defects  \
0                       guji-hambela  ...       Green
0
1                       guji-hambela  ...       Green
1
2                                NaN  ...         NaN
0
3                             oromia  ...       Green
2
4                       guji-hambela  ...       Green
2
...                              ...  ...         ...
...
1334               san juan, playas  ...  Blue-Green
1
1335               san juan, playas  ...  Blue-Green
0
1336  kwanza norte province, angola  ...         NaN
6
1337                             NaN  ...       Green
1
1338                             NaN  ...         NaN
9

               Expiration                    Certification.Body  \
0         April 3rd, 2016  METAD Agricultural Development plc
1         April 3rd, 2016  METAD Agricultural Development plc
2          May 31st, 2011         Specialty Coffee Association
3        March 25th, 2016  METAD Agricultural Development plc
4         April 3rd, 2016  METAD Agricultural Development plc
...                   ...                                  ...
1334   January 18th, 2017         Specialty Coffee Association
1335   January 18th, 2017         Specialty Coffee Association
1336  December 23rd, 2015         Specialty Coffee Association
1337    August 25th, 2015         Specialty Coffee Association
1338    August 25th, 2015         Specialty Coffee Association

                         Certification.Address  \
0     309fcf77415a3661ae83e027f7e5f05dad786e44
1     309fcf77415a3661ae83e027f7e5f05dad786e44
```

```
2     36d0d00a3724338ba7937c52a378d085f2172daa
3     309fcf77415a3661ae83e027f7e5f05dad786e44
4     309fcf77415a3661ae83e027f7e5f05dad786e44
...                                        ...
1334  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1335  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1336  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1337  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1338  ff7c18ad303d4b603ac3f8cff7e611ffc735e720

                          Certification.Contact unit_of_measurement  \
0     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
1     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
2     0878a7d4b9d35ddbf0fe2ce69a2062cceb45a660                    m
3     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
4     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
...                                        ...                  ...
1334  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1335  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1336  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1337  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1338  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m

     altitude_low_meters altitude_high_meters  altitude_mean_meters
0                 1950.0                2200.0                2075.0
1                 1950.0                2200.0                2075.0
2                 1600.0                1800.0                1700.0
3                 1800.0                2200.0                2000.0
4                 1950.0                2200.0                2075.0
...                  ...                   ...                   ...
1334                 NaN                   NaN                   NaN
1335                40.0                  40.0                  40.0
1336               795.0                 795.0                 795.0
1337                 NaN                   NaN                   NaN
1338                 NaN                   NaN                   NaN

[1339 rows x 43 columns]

b=a.drop(columns=['ICO.Number','Altitude','Harvest.Year','Grading.Date
','Expiration','Certification.Address','Certification.Contact'])

b

      Species                     Owner Country_of_Origin  \
0     Arabica                 metad plc          Ethiopia
1     Arabica                 metad plc          Ethiopia
2     Arabica  grounds for health admin         Guatemala
3     Arabica       yidnekachew dabessa          Ethiopia
4     Arabica                 metad plc          Ethiopia
...       ...                       ...               ...
```

```
1334  Robusta               luis robles           Ecuador
1335  Robusta               luis robles           Ecuador
1336  Robusta               james moore     United States
1337  Robusta             cafe politico             India
1338  Robusta             cafe politico           Vietnam

                                     Farm.Name Lot.Number
Mill  \
0                                    metad plc        NaN       metad
plc
1                                    metad plc        NaN       metad
plc
2     san marcos barrancas "san cristobal cuch        NaN
NaN
3        yidnekachew dabessa coffee plantation        NaN
wolensu
4                                    metad plc        NaN       metad
plc
...                                        ...        ...
...
1334                                 robustasa   Lavado 1     our own
lab
1335                                 robustasa   Lavado 3  own
laboratory
1336                           fazenda cazengo        NaN    cafe
cazengo
1337                                       NaN        NaN
NaN
1338                                       NaN        NaN
NaN

                                    Company
Region  \
0         metad agricultural developmet plc                   guji-
hambela
1         metad agricultural developmet plc                   guji-
hambela
2                                      NaN
NaN
3     yidnekachew debessa coffee plantation
oromia
4         metad agricultural developmet plc                   guji-
hambela
...                                    ...
...
1334                            robustasa               san juan,
playas
1335                            robustasa               san juan,
playas
```

```
1336                global opportunity fund  kwanza norte province,
angola
1337                           cafe politico
NaN
1338                           cafe politico
NaN

                                   Producer  Number_of_Bags  ...
Moisture  \
0                                  METAD PLC           300.0  ...
0.12
1                                  METAD PLC           300.0  ...
0.12
2                                        NaN             5.0  ...
0.00
3     Yidnekachew Dabessa Coffee Plantation           320.0  ...
0.11
4                                  METAD PLC           300.0  ...
0.12
...                                      ...             ...  ...
...
1334          Café Robusta del Ecuador S.A.             1.0  ...
0.00
1335          Café Robusta del Ecuador S.A.             1.0  ...
0.00
1336                           Cafe Cazengo             1.0  ...
0.00
1337                                     NaN             1.0  ...
0.10
1338                                     NaN             1.0  ...
0.12

     Category.One.Defects Quakers       Color Category.Two.Defects  \
0                       0     0.0       Green                   0
1                       0     0.0       Green                   1
2                       0     0.0         NaN                   0
3                       0     0.0       Green                   2
4                       0     0.0       Green                   2
...                   ...     ...         ...                 ...
1334                    0     0.0  Blue-Green                   1
1335                    0     0.0  Blue-Green                   0
1336                    0     0.0         NaN                   6
1337                   20     0.0       Green                   1
1338                   63     0.0         NaN                   9

                     Certification.Body  unit_of_measurement  \
0     METAD Agricultural Development plc                    m
1     METAD Agricultural Development plc                    m
2           Specialty Coffee Association                    m
3     METAD Agricultural Development plc                    m
```

```
4     METAD Agricultural Development plc                    m
...                                    ...                  ...
1334        Specialty Coffee Association                    m
1335        Specialty Coffee Association                    m
1336        Specialty Coffee Association                    m
1337        Specialty Coffee Association                    m
1338        Specialty Coffee Association                    m

      altitude_low_meters  altitude_high_meters  altitude_mean_meters

0                  1950.0                2200.0                2075.0

1                  1950.0                2200.0                2075.0

2                  1600.0                1800.0                1700.0

3                  1800.0                2200.0                2000.0

4                  1950.0                2200.0                2075.0

...                   ...                   ...                   ...

1334                NaN                   NaN                   NaN

1335               40.0                  40.0                  40.0

1336              795.0                 795.0                 795.0

1337                NaN                   NaN                   NaN

1338                NaN                   NaN                   NaN


[1339 rows x 36 columns]
```

```
#Returns the count of missing values in each column, helping identify
data quality issues.
b.isnull().sum()
```

```
Species                    0
Owner                      7
Country_of_Origin          1
Farm.Name                359
Lot.Number              1063
Mill                     318
Company                  209
Region                    59
Producer                 232
Number_of_Bags             1
Bag.Weight                 0
In.Country.Partner         0
```

```
Owner_1                   7
Variety                 226
ProcessingMethod        170
Aroma                     0
Flavor                    0
Aftertaste                0
Acidity                   0
Body                      0
Balance                   0
Uniformity                0
Clean.Cup                 0
Sweetness                 0
Cupper.Points             0
Total.Cup.Points          0
Moisture                  0
Category.One.Defects      0
Quakers                   1
Color                   270
Category.Two.Defects      0
Certification.Body        0
unit_of_measurement       0
altitude_low_meters     230
altitude_high_meters    230
altitude_mean_meters    230
dtype: int64
```

```
#Removing columns with more than 15% misssing value
c=b.drop(columns=['Farm.Name','Lot.Number','Mill','Company','Producer'
,'Variety','Color','altitude_low_meters','altitude_high_meters','altit
ude_mean_meters'])
```

```
c
```

```
      Species                    Owner Country_of_Origin  \
0     Arabica                metad plc          Ethiopia
1     Arabica                metad plc          Ethiopia
2     Arabica  grounds for health admin         Guatemala
3     Arabica       yidnekachew dabessa          Ethiopia
4     Arabica                metad plc          Ethiopia
...       ...                      ...                ...
1334  Robusta              luis robles           Ecuador
1335  Robusta              luis robles           Ecuador
1336  Robusta              james moore     United States
1337  Robusta            cafe politico             India
1338  Robusta            cafe politico           Vietnam

                           Region  Number_of_Bags Bag.Weight  \
0                    guji-hambela           300.0      60 kg
1                    guji-hambela           300.0      60 kg
2                             NaN             5.0          1
```

```
3                            oromia           320.0      60 kg
4                      guji-hambela           300.0      60 kg
...                             ...             ...        ...
1334               san juan, playas             1.0       2 kg
1335               san juan, playas             1.0       2 kg
1336  kwanza norte province, angola             1.0       1 kg
1337                            NaN             1.0      5 lbs
1338                            NaN             1.0      5 lbs

                      In.Country.Partner                    Owner_1  \
0     METAD Agricultural Development plc                  metad plc
1     METAD Agricultural Development plc                  metad plc
2           Specialty Coffee Association  Grounds for Health Admin
3     METAD Agricultural Development plc       Yidnekachew Dabessa
4     METAD Agricultural Development plc                  metad plc
...                                  ...                       ...
1334        Specialty Coffee Association                Luis Robles
1335        Specialty Coffee Association                Luis Robles
1336        Specialty Coffee Association                James Moore
1337        Specialty Coffee Association              Cafe Politico
1338        Specialty Coffee Association              Cafe Politico

     ProcessingMethod  Aroma  ...  Clean.Cup  Sweetness  Cupper.Points
\
0        Washed / Wet   8.67  ...      10.00      10.00           8.75

1        Washed / Wet   8.75  ...      10.00      10.00           8.58

2                 NaN   8.42  ...      10.00      10.00           9.25

3       Natural / Dry   8.17  ...      10.00      10.00           8.67

4        Washed / Wet   8.25  ...      10.00      10.00           8.58

...               ...    ...  ...        ...        ...            ...

1334              NaN   7.75  ...      10.00       7.75           7.83

1335              NaN   7.50  ...      10.00       8.42           8.58

1336    Natural / Dry   7.33  ...       9.33       7.42           7.17

1337    Natural / Dry   7.42  ...       9.33       7.08           6.92

1338    Natural / Dry   6.75  ...       9.33       6.67           7.92


      Total.Cup.Points  Moisture  Category.One.Defects  Quakers  \
0                90.58      0.12                     0      0.0
1                89.92      0.12                     0      0.0
2                89.75      0.00                     0      0.0
```

```
3                89.00      0.11                       0      0.0
4                88.83      0.12                       0      0.0
...                ...       ...                     ...      ...
1334             78.75      0.00                       0      0.0
1335             78.08      0.00                       0      0.0
1336             77.17      0.00                       0      0.0
1337             75.08      0.10                      20      0.0
1338             73.75      0.12                      63      0.0

      Category.Two.Defects                  Certification.Body  \
0                        0  METAD Agricultural Development plc
1                        1  METAD Agricultural Development plc
2                        0        Specialty Coffee Association
3                        2  METAD Agricultural Development plc
4                        2  METAD Agricultural Development plc
...                    ...                                 ...
1334                     1        Specialty Coffee Association
1335                     0        Specialty Coffee Association
1336                     6        Specialty Coffee Association
1337                     1        Specialty Coffee Association
1338                     9        Specialty Coffee Association

      unit_of_measurement
0                       m
1                       m
2                       m
3                       m
4                       m
...                   ...
1334                    m
1335                    m
1336                    m
1337                    m
1338                    m

[1339 rows x 26 columns]
```

```
c.isnull().sum()
```

```
Species                    0
Owner                      7
Country_of_Origin          1
Region                    59
Number_of_Bags             1
Bag.Weight                 0
In.Country.Partner         0
Owner_1                    7
ProcessingMethod         170
Aroma                      0
Flavor                     0
```

```
Aftertaste                0
Acidity                   0
Body                      0
Balance                   0
Uniformity                0
Clean.Cup                 0
Sweetness                 0
Cupper.Points             0
Total.Cup.Points          0
Moisture                  0
Category.One.Defects      0
Quakers                   1
Category.Two.Defects      0
Certification.Body        0
unit_of_measurement       0
dtype: int64
```

```
c.describe()
```

```
       Number_of_Bags        Aroma       Flavor   Aftertaste
Acidity  \
count     1338.000000  1339.000000  1339.000000  1339.000000
1339.000000
mean       159.085202     7.770187     7.520426     7.401083
7.535706
std        173.698167     5.534440     0.398442     0.404463
0.379827
min          0.000000     0.000000     0.000000     0.000000
0.000000
25%         14.000000     7.420000     7.330000     7.250000
7.330000
50%        175.000000     7.580000     7.580000     7.420000
7.580000
75%        275.000000     7.750000     7.750000     7.580000
7.750000
max       3200.000000   200.000000     8.830000     8.670000
8.750000

              Body      Balance   Uniformity    Clean.Cup    Sweetness
\
count  1339.000000  1339.000000  1339.000000  1339.000000  1339.000000

mean      7.517498     7.518013     9.834877     9.835108     9.856692

std       0.370064     0.408943     0.554591     0.763946     0.616102

min       0.000000     0.000000     0.000000     0.000000     0.000000

25%       7.330000     7.330000    10.000000    10.000000    10.000000
```

```
50%       7.500000     7.500000    10.000000    10.000000    10.000000

75%       7.670000     7.750000    10.000000    10.000000    10.000000

max       8.580000     8.750000    10.000000    10.000000    10.000000


       Cupper.Points  Total.Cup.Points     Moisture
Category.One.Defects  \
count    1339.000000       1339.000000  1339.000000
1339.000000
mean        7.503376         82.089851     0.088379
0.479462
std         0.473464          3.500575     0.048287
2.549683
min         0.000000          0.000000     0.000000
0.000000
25%         7.250000         81.080000     0.090000
0.000000
50%         7.500000         82.500000     0.110000
0.000000
75%         7.750000         83.670000     0.120000
0.000000
max        10.000000         90.580000     0.280000
63.000000

           Quakers  Category.Two.Defects
count  1338.000000           1339.000000
mean      0.173393              3.556385
std       0.832121              5.312541
min       0.000000              0.000000
25%       0.000000              0.000000
50%       0.000000              2.000000
75%       0.000000              4.000000
max      11.000000             55.000000
```

```
#Checking for all the unique values
c['Owner'].unique
```

```
<bound method Series.unique of 0                       metad plc
1                       metad plc
2       grounds for health admin
3            yidnekachew dabessa
4                       metad plc
                  ...
1334                 luis robles
1335                 luis robles
1336                 james moore
1337               cafe politico
```

```
1338                cafe politico
Name: Owner, Length: 1339, dtype: object>
```

```
#Getting the mode for the column & storing it in a variable
Owner=c.Owner.mode()[0]
Owner
```

```
'juan luis alvarado romero'
```

```
c.Owner.fillna(Owner,inplace=True)
c.isnull().sum()
```

```
Species                    0
Owner                      0
Country_of_Origin          1
Region                    59
Number_of_Bags             1
Bag.Weight                 0
In.Country.Partner         0
Owner_1                    7
ProcessingMethod         170
Aroma                      0
Flavor                     0
Aftertaste                 0
Acidity                    0
Body                       0
Balance                    0
Uniformity                 0
Clean.Cup                  0
Sweetness                  0
Cupper.Points              0
Total.Cup.Points           0
Moisture                   0
Category.One.Defects       0
Quakers                    1
Category.Two.Defects       0
Certification.Body         0
unit_of_measurement        0
dtype: int64
```

```
c['Country_of_Origin'].unique
```

```
<bound method Series.unique of 0             Ethiopia
1              Ethiopia
2             Guatemala
3              Ethiopia
4              Ethiopia
              ...
1334           Ecuador
1335           Ecuador
1336    United States
```

```
1337            India
1338          Vietnam
Name: Country_of_Origin, Length: 1339, dtype: object>
```

```
Country_of_Origin =c.Country_of_Origin.mode()[0]
Country_of_Origin
```

```
'Mexico'
```

```
c.Country_of_Origin.fillna(Country_of_Origin,inplace=True)
c.isnull().sum()
```

```
Species                    0
Owner                      0
Country_of_Origin          0
Region                    59
Number_of_Bags             1
Bag.Weight                 0
In.Country.Partner         0
Owner_1                    7
ProcessingMethod         170
Aroma                      0
Flavor                     0
Aftertaste                 0
Acidity                    0
Body                       0
Balance                    0
Uniformity                 0
Clean.Cup                  0
Sweetness                  0
Cupper.Points              0
Total.Cup.Points           0
Moisture                   0
Category.One.Defects       0
Quakers                    1
Category.Two.Defects       0
Certification.Body         0
unit_of_measurement        0
dtype: int64
```

```
c['Region'].unique
```

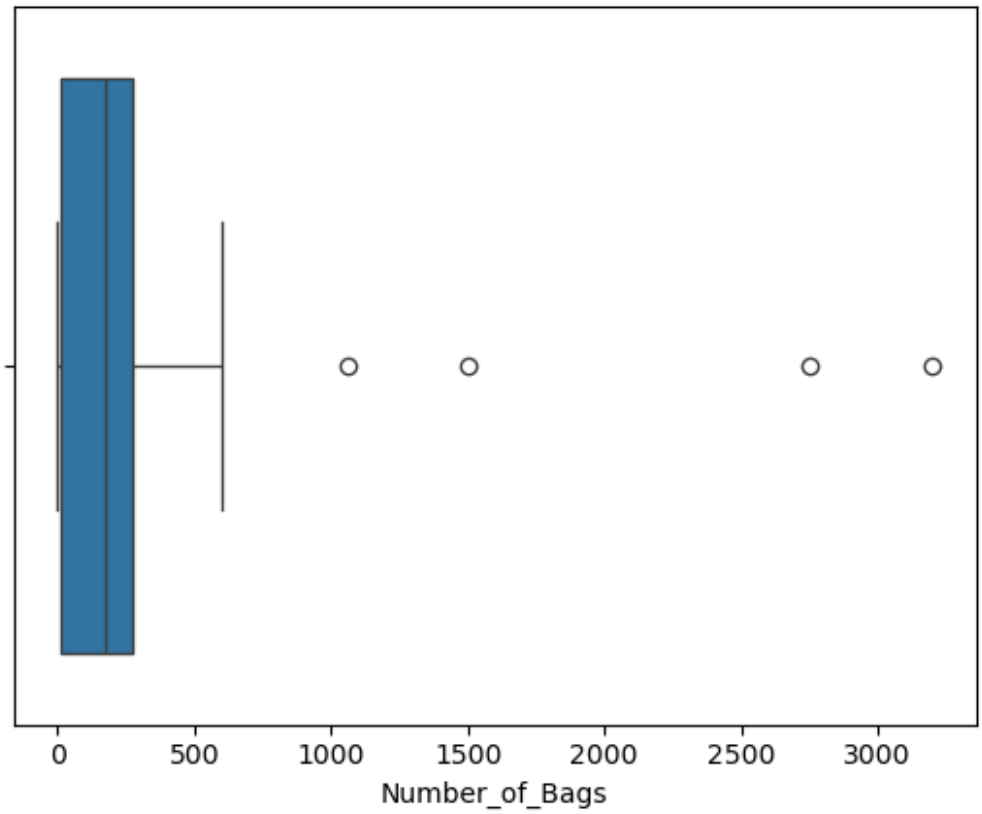
```
<bound method Series.unique of 0                          guji-hambela
1                        guji-hambela
2                                 NaN
3                              oromia
4                        guji-hambela
                     ...
1334                  san juan, playas
1335                  san juan, playas
1336    kwanza norte province, angola
```

```
1337                                NaN
1338                                NaN
Name: Region, Length: 1339, dtype: object>

Region=c.Region.mode()[0]
Region

'huila'

c.Region.fillna(Region,inplace=True)
c.isnull().sum()

Species                     0
Owner                       0
Country_of_Origin           0
Region                      0
Number_of_Bags              1
Bag.Weight                  0
In.Country.Partner          0
Owner_1                     7
ProcessingMethod          170
Aroma                       0
Flavor                      0
Aftertaste                  0
Acidity                     0
Body                        0
Balance                     0
Uniformity                  0
Clean.Cup                   0
Sweetness                   0
Cupper.Points               0
Total.Cup.Points            0
Moisture                    0
Category.One.Defects        0
Quakers                     1
Category.Two.Defects        0
Certification.Body          0
unit_of_measurement         0
dtype: int64

c['Number_of_Bags'].unique()

array([3.000e+02, 5.000e+00, 3.200e+02, 1.000e+02,       nan,
5.000e+01,
       1.000e+01, 1.000e+00, 1.500e+02, 3.000e+00, 2.500e+02,
1.400e+01,
       2.750e+02, 2.000e+01, 2.900e+01, 2.500e+01, 5.300e+01,
1.200e+01,
       7.000e+00, 8.000e+01, 3.700e+01, 2.800e+02, 1.900e+01,
8.000e+00,
       1.600e+01, 2.000e+00, 3.600e+01, 3.600e+02, 5.400e+01,
```

```
1.300e+01,
       2.700e+01, 2.000e+02, 1.350e+02, 1.700e+02, 3.800e+01,
3.100e+01,
       1.500e+01, 2.430e+02, 2.520e+02, 1.340e+02, 4.000e+00,
1.200e+02,
       2.750e+03, 2.350e+02, 1.250e+02, 6.600e+01, 7.500e+01,
1.100e+01,
       3.500e+01, 5.600e+01, 3.040e+02, 6.900e+01, 1.500e+03,
2.300e+02,
       2.480e+02, 6.500e+01, 3.770e+02, 1.300e+02, 3.050e+02,
3.200e+03,
       1.380e+02, 2.700e+02, 4.500e+01, 2.260e+02, 4.800e+01,
1.670e+02,
       1.750e+02, 1.800e+01, 2.850e+02, 3.300e+01, 2.450e+02,
1.800e+02,
       6.000e+02, 5.000e+02, 3.900e+01, 6.000e+00, 2.200e+02,
2.600e+01,
       3.000e+01, 2.320e+02, 8.400e+01, 9.000e+01, 3.100e+02,
3.250e+02,
       1.700e+01, 1.210e+02, 2.300e+01, 1.290e+02, 4.000e+01,
3.200e+01,
       2.100e+01, 6.000e+01, 9.300e+01, 7.700e+01, 2.880e+02,
1.980e+02,
       7.000e+01, 4.200e+01, 2.800e+01, 4.300e+01, 4.900e+01,
7.400e+01,
       5.100e+01, 0.000e+00, 4.400e+01, 1.062e+03, 1.490e+02,
2.740e+02,
       1.140e+02, 4.500e+02, 6.200e+01, 1.660e+02, 2.400e+01,
3.020e+02,
       5.800e+01, 1.650e+02, 5.500e+02, 1.230e+02, 2.400e+02,
1.600e+02,
       9.400e+01, 4.400e+02, 2.200e+01, 2.560e+02, 4.000e+02,
8.200e+01,
       2.090e+02, 3.800e+02, 2.530e+02, 2.230e+02, 1.270e+02,
2.020e+02,
       9.000e+00, 8.500e+01, 1.400e+02])
```

```
#Plotting a boxplot for the column
import seaborn as sns
sns.boxplot(x=df['Number_of_Bags'])
```

```
<Axes: xlabel='Number_of_Bags'>
```




```
Number_of_Bags=c['Number_of_Bags'].median()

#Sorting the values in ascending order to find median
c['Number_of_Bags'].sort_values(ascending=True).head()
```

```
704     0.0
1206    1.0
379     1.0
1188    1.0
444     1.0
Name: Number_of_Bags, dtype: float64
```

```
c.Number_of_Bags.fillna(Number_of_Bags,inplace=True)
c.isnull().sum()
```

```
Species                   0
Owner                     0
Country_of_Origin         0
Region                    0
Number_of_Bags            0
Bag.Weight                0
In.Country.Partner        0
Owner_1                   7
ProcessingMethod        170
Aroma                     0
```

```
Flavor                    0
Aftertaste                0
Acidity                   0
Body                      0
Balance                   0
Uniformity                0
Clean.Cup                 0
Sweetness                 0
Cupper.Points             0
Total.Cup.Points          0
Moisture                  0
Category.One.Defects      0
Quakers                   1
Category.Two.Defects      0
Certification.Body        0
unit_of_measurement       0
dtype: int64
```

```
c['ProcessingMethod'].unique()
```

```
array(['Washed / Wet', nan, 'Natural / Dry', 'Pulped natural / honey',
       'Semi-washed / Semi-pulped', 'Other'], dtype=object)
```

```
ProcessingMethod=c.ProcessingMethod.mode()[0]
ProcessingMethod
```

```
'Washed / Wet'
```

```
c.ProcessingMethod.fillna(ProcessingMethod,inplace=True)
c.isnull().sum()
```

```
Species                 0
Owner                   0
Country_of_Origin       0
Region                  0
Number_of_Bags          0
Bag.Weight              0
In.Country.Partner      0
Owner_1                 7
ProcessingMethod        0
Aroma                   0
Flavor                  0
Aftertaste              0
Acidity                 0
Body                    0
Balance                 0
Uniformity              0
Clean.Cup               0
Sweetness               0
Cupper.Points           0
Total.Cup.Points        0
```

```
Moisture                0
Category.One.Defects    0
Quakers                 1
Category.Two.Defects    0
Certification.Body      0
unit_of_measurement     0
dtype: int64
```

```
c['Owner_1'].unique()
```

```
array(['metad plc', 'Grounds for Health Admin', 'Yidnekachew Dabessa',
       'Ji-Ae Ahn', 'Hugo Valdivia', 'Ethiopia Commodity Exchange',
       'Diamond Enterprise Plc', 'Mohammed Lalo',
       'CQI Q Coffee Sample Representative', 'Yunnan Coffee Exchange',
       'EssenceCoffee', 'The Coffee Source Inc.', 'ROBERTO LICONA
FRANCO',
       'NUCOFFEE', 'Kabum Trading company', 'Bismarck Castro',
       'Lin, Che-Hao Krude 林哲豪', 'Nora Zeas', 'Specialty Coffee-
Korea',
       'Francisco A Mena', 'Hider Abamecha', 'Daniel Magu',
       'Kona Pacific Farmers Cooperative', 'ITDP International',
       'Jacques Pereira Carneiro', 'Jungle Estate',
       'Great Lakes Coffee Uganda', 'LUSSO LAB', 'AFCA',
       'Juan Luis Alvarado Romero', 'Kawacom Uganda LTD',
       'Exportadora de Cafe Condor S.A', 'Gonzalo Hernandez',
       'Ibrahim Hussien Speciality Coffee Producer &Export',
       'SEID DAMTEW COFFEE PLANATAION', 'Dane Loraas',
       'Colbran Coffeelands, Ltd.', 'Atlantic Specialty Coffee',
       'Assefa Belay Coffee Producer', 'Kyagalanyi Ltd',
       'RASHID MOLEDINA & CO. (MSA) LTD.', 'Ibero Kenya Limited',
       'Compañia Colombiana Agroindustrial S.A',
       'Nomura Trading Co., Ltd.', 'CARCAFE LTDA CI', 'Steven Kil',
       'Eileen Koyanagi', 'Kyagalanyi Coffee Ltd', 'Racafe & Cia
S.C.A',
       'Troy Quimby', 'El Equimite, Cafetal Biodinámico', 'SIMON
MAHINDA',
       'Young Kim', 'Carl Walker', 'Taylor Winch (T) Ltd',
       'ARTEMIO ZAPATA TEJEDA', 'Brian Speckman', 'Philip Schluter',
       '松澤宏樹Koju Matsuzawa', 'Lydiah Mwangi', 'CADEXSA',
       'Consejo Salvadoreño del Café', 'SanJava Coffee', 'Rodrigo
Soto',
       'Fabian Calderon Mora', 'Eric Thormaehlen', 'Rob Tuttle',
       'CQI Taiwan ICP CQI台灣合作夥伴', 'Dream Together',
       'ORGANIZACIONES DE PRODUCTORES DE CAFE COLIMENSE',
       'Benjamin Schmerler', 'Taylor Winch (Coffee) Ltd.', 'Max
Gurdian',
       'ECOM Japan Limited', 'Federacion Nacional de Cafeteros',
       'Eric Wu', 'MARIA IMELDA USCANGA MARTINEZ', 'ALFREDO BOJALIL',
       'Daniel Friedlander', 'Alexandra Katona-Carroll',
       'Aulia Arif Syahri', 'Kao Ming Lee',
```

```
       'MARIA AMALIA GUADALUPE TORIELLO ELORZA', 'Raúl Vargas',
       'VICTOR HUGO MELCHOR CORDOVA', 'Tembo Coffee Company Ltd',
       'JESUS SALAZAR VELASCO', 'MANUEL HERRERA JUAREZ', 'Wayner
Jimenez',
       'COOPERATIVA EL GORRION R.L', 'Cafebras', 'CECA,S.A.',
       'Asefa Dukamo Keroma', 'Selian Coffee Estate',
       'Olam Agro Colombia', 'Chris Finch', 'ITOCHU Corporation',
       'Owen Carver', 'PT.ROYAL PACIFIC INDAH INTERNATIONAL',
       'ANDRES MARTINEZ LEON', 'Amanda Powers', 'Ipanema Coffees',
       'Doi Tung Development Project', 'CAFES TOMARI SA DE CV',
       'Sarawut Premjit', 'ALMACAFE', 'OSCAR ORTEGA CARBALLO',
       'CECA, S.A.', 'yasmin Cofffee Plantation Plc', 'Garet Alban',
       'FILEMON MENDOZA CAMPOS', 'Doi Chaang Coffee Company',
       'Kennedy Macharia', 'Nile Highland Arabica Coffee Farmers',
       'German Negron', 'SAUL M. HERNANDEZ RAMIREZ',
       'COMERCIAL INTERNACIONAL EXPORTADORA, S.A.', 'Rob Stephen',
       'JUAN LUIS ORTEGA CARBALLO', 'EKAI International Company Ltd.',
       'ANDREAS KUSSMAUL', 'Bulamburi coffee farmers association',
       'Damari Absalome', 'Debesa Agro Industry Plc', nan,
       'MIGUEL CORTES MORENO', 'GABRIEL BERNARDO RIVAS ROSS',
       'Felipe Isaza', 'Specialty Coffee Association of Indonesia',
       'Bugisu Cooperative Union', 'BOURBON SPECIALTY COFFEES',
       'Ngila Estate Ltd', 'Federación Nacional de Cafeteros',
       'J.ANDRADE', 'ITIAH COFFEE LLC',
       'CAFE DE DON BALBINO S.C. DE R.L. DE C.V.',
       'PRODUCTOS Y SERVICIOS CHILINDRON S.A. DE C.V.',
       'CALIXTO GUILLEN VAZQUEZ', 'ERNESTO RODRIGUEZ LUNA',
       'MODESTO LANDEROS FLORES', 'ANDREA BERNAL', 'Sunvirtue Co.,
Ltd.',
       'Tutunze Kahawa Ltd', 'Cafe Politico', 'Mayra Yessenia Torres',
       'Balam Hinyula', 'NESTOR MENDEZ GOMEZ',
       'FERNANDO MENDOZA APARICIO',
       'MARIA LUISA DEL CARMEN ROJAS NARVAEZ', 'UCFA',
       'Irene Alves Santos', 'Star Cafe Ltd',
       'ROSA AURORA FALCON FERNANDEZ', 'SANTIAGO SOLIS AYERDI',
       'Renee A. Perrine', 'Zarah Zamora Perez', 'Andrew Bowman',
       'Expocaccer Coop dos Cafeic do Cerrado Ltda',
       'Nyapea coffee farmers association', 'MARIA GUADALUPE GOMEZ
ANZO',
       'Royal Base Corporation', 'VERONICA LOPEZ CASTILLEJOS',
       'Samuel Muhirwa', 'Joshua Marsceau', 'Coffeebythebag.com ,
INC',
       'Edwin Agasso', 'ARMANDO LUIS POHLENZ MARTINEZ', 'Coffee
Export',
       'SERGIO DE LA VEQUIA BERNARDI', 'ROMULO BELLO FLORES',
       'Rachel Peterson', 'José Luis Rojas Yeo', 'Nitin Coffee
Estate',
       'Adam Kline', 'MONTEGRANDE',
       'GRUPO CAFETALERO LOS BRUJOS SPR DE RL', 'George A. Fernandez',
```

```
       'Gabriel Barbara', 'Andry Simarmata', 'Brent Hall',
       'GUILLERMO ROJAS SALDANA', 'Elsy Reyes', 'Shah Plantations',
       'Amkeni Gourmet Coffee Group', 'ENRIQUE MITRE LOPEZ',
       'Enrique Eduardo Lopez Aguilar', 'Brian Beck',
       'Gladness Obed Pallangyo', 'DARIO CESAR GALEANA SANCHEZ',
       'JOSE DANIEL COBILT CASTRO', 'ALVARO QUIROS PEREZ',
       'OLIVIA HERNANDEZ VIRVES', 'FINCA LAS NIEVES',
       'Pedro Santos e Silva', 'Michael Gavina', 'KlemOrganics',
       'JESUS CARLOS CARDENAS VALDIVIA', 'BENCAFE, S. A.',
       'Langiro Farm group', 'IBERO COFFEE TRADING CO (T) LTD',
       'SALVADOR CARO CARRION', 'CAFETALERA INTERNACIONAL CAFINTER,
S.A.',
       'Ngorogoro Convenant Estate', 'JULIO PEREZ HERNANDEZ', 'Didas',
       'Minwook Ku', 'Finca Estate', 'Beneficio Santa Rosa',
       'JORGE OCTAVIO ESCAMILLA PRADO', 'Mcomafa Co Ltd',
       'JUAN HERMILIO SAMPIERI CARCAMO', 'U Mg Mg', 'VIRIDIANA',
       'Kurt Kappeli', 'CHRISTINA DUSING',
       'JORGE FRANCISCO MARTINEZ HACHITY', 'SERGIO LANDA ALARCON',
       'DIEGO MANUEL WOOLRICH RAMIREZ', 'DAE Ltd Company',
       'FREDY GORDILLO REYES', 'VIRGINIA GORDILLO GORDILLO',
       'JOSE LUIS MUNOZ GUERRERO', 'MDH', 'Acacia Hills Ltd',
       'Exportadora Atlantic, S.A.', 'Genius Coffee',
       'Santa Laura Exportadora de Cafe S.L.E.C. S.A.',
       'Lin, Che-Hao Krude 林哲豪\n', 'Myriam Kaplan-Pasternak',
       'TOMAS EDELMANN BLASS', 'MARIA DE LA PAZ AGUILAR GUILLEN',
       'Angel Oscar Medina Rodriguez', 'Victoria',
       'HECTOR GABRIEL BARREDA NADER', 'Shangrilla Estate Ltd',
       'Immaculata John', 'KERCHANSHE', 'Gregorio Sebba',
       'Rolando Lacayo', 'Wali Ali', 'OBED RENDON PONCE',
       'GERARDO HERNANDEZ VALDERRABANO', 'BALBINO RAMIREZ FLORES',
       'Mlimani Ngarashi', 'ALEJANDRO GARCIA PALACIOS',
       'Grupo Santab S.A de C.V.', 'Min Hlaing', 'Karatu Estate',
       'EDUARDO LUIS AUGUSTO VELAZQUEZ SOLIS',
       'LUIS ROBERTO FERMOSO BELTRAN', 'JOSE MANUEL VERGARA CORTES',
       'U Soe', 'Burka Coffee Estate', 'Janny Marlith Torres',
       'Case Noyale Ltd', 'Shwe Yin Mar Coffee',
       'ISRAEL EDUARDO PAZ GARCIA', 'Adam Ciruli Ye',
       'CQI Taiwan ICP CQI台灣合作夥伴\n', 'Delfina Leon Shine',
       'Kongoni Estate', 'Volcafe Ltda. - Brasil', 'Bob McCauley',
       'U Htun Htun', 'Gloria Antonieta Escobar Urrutia',
       'Honor dela Fuente', 'PABLO ENRIQUE MARTINEZ GAMA',
       'MARCO VIRGILIO RAMIREZ TELIZ', 'Brayan Cunha Souza',
       'FEDERICO PACHECO PEREZ', 'Ngu Shwe Li',
       'SEMIRAMIS CASAS VELAZQUEZ', 'JESUS CARLOS CADENA VALDIVIA',
       'Asociación Aldea Global Jinotega', 'LEONIDES DE LA CRUZ
LOPEZ',
       'MARIO JOSE FERNANDEZ', 'ADRIANA TORRES RICO QUEVEDO',
       'Rre Kunene', 'ERIC JESUS CORDOBA ARROYO',
       'JULIO CESAR ROBLES FLORES', 'Masamichi Hiroike',
```

```
       'JUANA RODRIGUEZ GUTIERREZ',
       'CAFES FINOS DE EXPORTACION S DE R.L.',
       'Sustainable Harvest Coffee', 'GONZALO DE AQUINO FLORES',
       'JUAN AVENAMAR RODRIGUEZ FUNEZ', 'OCTAVIO AUGUSTO DIAZ TREJO',
       'DAMASO MARTINEZ PEREZ',
       'PRODUCTORES DE ESPECIALIDAD EMILIANO ZAPEATA, SPR.',
       'JUAN GARCIA HERNANDEZ', 'ROSARIO MIGUEL HERNANDEZ',
       'FRANCISCO RUIZ NUNEZ', 'PABLO CERVANTES MORELOS         ',
       'GUSTAVO AMIEVA GONZALEZ', 'Samuel Eli Gurel', 'Mao-Heng Chu',
       'GUSTAVO ABARCA SOLIS', 'STEPHANY ESCAMILLA FEMAT',
       'HOMERO ANTONIO DE ANDA ANDRADE', 'William Ho',
       'GUILLERMO EDUARDO BOBADILLA MUGUIRA', 'Ana Gonzales',
       'FRANCISCO HERNANDEZ LORENZO', 'MARTIN JIMENEZ CASIANO',
       'GRUPO JUVENIL MAGTAYANI, AC', 'MYRNA ROXANA GALVEZ GONZALEZ',
       'EUGENE HOLMAN PEW', 'JOSE ARMANDO NORBERTO BORZANI LEMINI',
       'RICARDO AARON SAMPIERI MARINI', 'JUAN CARLOS GARCIA LOPEZ',
       'Ankole coffee producers coop', 'Nishant Gurjer', 'Andrew
Hetzel',
       'UGACOF', 'Katuka Development Trust Ltd',
       'Kasozi Coffee Farmers Association', 'Nitubaasa Ltd',
       'Mannya coffee project', 'Luis Robles', 'James Moore'],
      dtype=object)
```

```
Owner_1=c.Owner_1.mode()[0]
Owner_1
```

```
'Juan Luis Alvarado Romero'
```

```
c.Owner_1.fillna(Owner_1,inplace=True)
c.isnull().sum()
```

```
Species                 0
Owner                   0
Country_of_Origin       0
Region                  0
Number_of_Bags          0
Bag.Weight              0
In.Country.Partner      0
Owner_1                 0
ProcessingMethod        0
Aroma                   0
Flavor                  0
Aftertaste              0
Acidity                 0
Body                    0
Balance                 0
Uniformity              0
Clean.Cup               0
Sweetness               0
Cupper.Points           0
```

```
Total.Cup.Points         0
Moisture                 0
Category.One.Defects     0
Quakers                  1
Category.Two.Defects     0
Certification.Body       0
unit_of_measurement      0
dtype: int64
```

```
c['Quakers'].unique()
```

```
array([ 0.,  1.,  4.,  2.,  5.,  6.,  3., 11.,  7., nan,  9.,  8.])
```

```
c.Quakers.mode()[0]
```

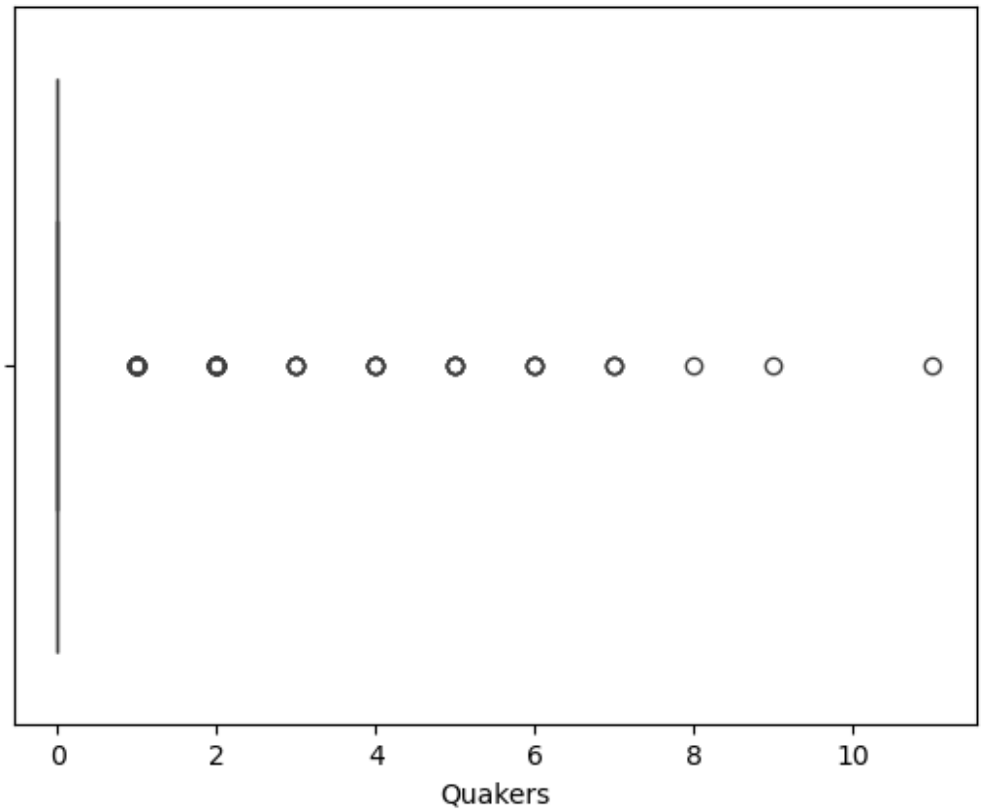
```
0.0
```

```
c['Quakers'].dtype
```

```
dtype('float64')
```

```
sns.boxplot(x=df['Quakers'])
```

```
<Axes: xlabel='Quakers'>
```



```
c['Quakers'].sort_values(ascending=True)
```

```
0         0.0
892       0.0
891       0.0
890       0.0
889       0.0
         ...
1260      7.0
1186      8.0
637       9.0
241      11.0
366       NaN
Name: Quakers, Length: 1339, dtype: float64
```

```
Quakers=df.Quakers.median()
Quakers
```

```
0.0
```

```
c.Quakers.fillna(Quakers,inplace=True)
c.isnull().sum()
```

```
Species                   0
Owner                     0
Country_of_Origin         0
Region                    0
Number_of_Bags            0
Bag.Weight                0
In.Country.Partner        0
Owner_1                   0
ProcessingMethod          0
Aroma                     0
Flavor                    0
Aftertaste                0
Acidity                   0
Body                      0
Balance                   0
Uniformity                0
Clean.Cup                 0
Sweetness                 0
Cupper.Points             0
Total.Cup.Points          0
Moisture                  0
Category.One.Defects      0
Quakers                   0
Category.Two.Defects      0
Certification.Body        0
unit_of_measurement       0
dtype: int64
```

```
c.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1339 entries, 0 to 1338
Data columns (total 26 columns):
 #   Column                Non-Null Count  Dtype  
---  ------                --------------  -----  
 0   Species               1339 non-null   object 
 1   Owner                 1339 non-null   object 
 2   Country_of_Origin     1339 non-null   object 
 3   Region                1339 non-null   object 
 4   Number_of_Bags        1339 non-null   float64
 5   Bag.Weight            1339 non-null   object 
 6   In.Country.Partner    1339 non-null   object 
 7   Owner_1               1339 non-null   object 
 8   ProcessingMethod      1339 non-null   object 
 9   Aroma                 1339 non-null   float64
 10  Flavor                1339 non-null   float64
 11  Aftertaste            1339 non-null   float64
 12  Acidity               1339 non-null   float64
 13  Body                  1339 non-null   float64
 14  Balance               1339 non-null   float64
 15  Uniformity            1339 non-null   float64
 16  Clean.Cup             1339 non-null   float64
 17  Sweetness             1339 non-null   float64
 18  Cupper.Points         1339 non-null   float64
 19  Total.Cup.Points      1339 non-null   float64
 20  Moisture              1339 non-null   float64
 21  Category.One.Defects  1339 non-null   int64  
 22  Quakers               1339 non-null   float64
 23  Category.Two.Defects  1339 non-null   int64  
 24  Certification.Body    1339 non-null   object 
 25  unit_of_measurement   1339 non-null   object 
dtypes: float64(14), int64(2), object(10)
memory usage: 272.1+ KB
```

```
#Removing all the objects to only have numerical columns.
df1 =c.select_dtypes(exclude=['object'])
df1
```

```
      Number_of_Bags  Aroma  Flavor  Aftertaste  Acidity  Body  
Balance  \
0              300.0   8.67    8.83        8.67     8.75  8.50  
8.42
1              300.0   8.75    8.67        8.50     8.58  8.42  
8.42
2                5.0   8.42    8.50        8.42     8.42  8.33  
8.42
3              320.0   8.17    8.58        8.42     8.42  8.50  
8.25
```

```
4              300.0   8.25    8.50        8.25     8.50  8.42
8.33
...              ...    ...     ...         ...      ...   ...      ..
.
1334             1.0   7.75    7.58        7.33     7.58  5.08
7.83
1335             1.0   7.50    7.67        7.75     7.75  5.17
5.25
1336             1.0   7.33    7.33        7.17     7.42  7.50
7.17
1337             1.0   7.42    6.83        6.75     7.17  7.25
7.00
1338             1.0   6.75    6.67        6.50     6.83  6.92
6.83

      Uniformity  Clean.Cup  Sweetness  Cupper.Points
Total.Cup.Points  \
0           10.00      10.00      10.00           8.75
90.58
1           10.00      10.00      10.00           8.58
89.92
2           10.00      10.00      10.00           9.25
89.75
3           10.00      10.00      10.00           8.67
89.00
4           10.00      10.00      10.00           8.58
88.83
...           ...        ...        ...            ...                ..
.
1334        10.00      10.00       7.75           7.83
78.75
1335        10.00      10.00       8.42           8.58
78.08
1336         9.33       9.33       7.42           7.17
77.17
1337         9.33       9.33       7.08           6.92
75.08
1338         9.33       9.33       6.67           7.92
73.75

      Moisture  Category.One.Defects  Quakers  Category.Two.Defects
0         0.12                     0      0.0                    0
1         0.12                     0      0.0                    1
2         0.00                     0      0.0                    0
3         0.11                     0      0.0                    2
4         0.12                     0      0.0                    2
...        ...                   ...      ...                  ...
1334      0.00                     0      0.0                    1
1335      0.00                     0      0.0                    0
```

```
1336      0.00                     0      0.0                     6
1337      0.10                    20      0.0                     1
1338      0.12                    63      0.0                     9

[1339 rows x 16 columns]
```

```
q1=df1.quantile(0.25)
q3=df1.quantile(0.75)
```

```
q1
```

```
Number_of_Bags           14.00
Aroma                     7.42
Flavor                    7.33
Aftertaste                7.25
Acidity                   7.33
Body                      7.33
Balance                   7.33
Uniformity               10.00
Clean.Cup                10.00
Sweetness                10.00
Cupper.Points             7.25
Total.Cup.Points         81.08
Moisture                  0.09
Category.One.Defects      0.00
Quakers                   0.00
Category.Two.Defects      0.00
Name: 0.25, dtype: float64
```

```
q3
```

```
Number_of_Bags          275.00
Aroma                     7.75
Flavor                    7.75
Aftertaste                7.58
Acidity                   7.75
Body                      7.67
Balance                   7.75
Uniformity               10.00
Clean.Cup                10.00
Sweetness                10.00
Cupper.Points             7.75
Total.Cup.Points         83.67
Moisture                  0.12
Category.One.Defects      0.00
Quakers                   0.00
Category.Two.Defects      4.00
Name: 0.75, dtype: float64
```

```
iqr=q3-q1
iqr
```

```
Number_of_Bags           261.00
Aroma                      0.33
Flavor                     0.42
Aftertaste                 0.33
Acidity                    0.42
Body                       0.34
Balance                    0.42
Uniformity                 0.00
Clean.Cup                  0.00
Sweetness                  0.00
Cupper.Points              0.50
Total.Cup.Points           2.59
Moisture                   0.03
Category.One.Defects       0.00
Quakers                    0.00
Category.Two.Defects       4.00
dtype: float64
```

```
#Finding the outliers & storing it in a variable
d=(df1<(q1-1.5*iqr))|(df1>(q3+1.5*iqr))
d.sum()
```

```
Number_of_Bags             4
Aroma                     75
Flavor                    44
Aftertaste                87
Acidity                   25
Body                      34
Balance                   40
Uniformity               187
Clean.Cup                120
Sweetness                121
Cupper.Points             33
Total.Cup.Points          72
Moisture                 305
Category.One.Defects     202
Quakers                   94
Category.Two.Defects      94
dtype: int64
```

```
#Removing the outliers from the original data frame
filter=c[~((df1<(q1-1.5*iqr))|(df1>(q3+1.5*iqr))).any(axis=1)]
filter
```

```
      Species                            Owner Country_of_Origin  \
21    Arabica          the coffee source inc.        Costa Rica
30    Arabica                      nora zeas         Nicaragua
34    Arabica                hider abamecha          Ethiopia
35    Arabica                   daniel magu             Kenya
43    Arabica          lin, che-hao krude 林哲豪    Taiwan
```

```
...       ...                                  ...                ...
1167  Arabica  exportadora de cafe condor s.a          Colombia
1182  Arabica          lin, che-hao krude 林哲豪     Taiwan
1183  Arabica    marco virgilio ramirez teliz             Mexico
1205  Arabica       israel eduardo paz garcia             Mexico
1209  Arabica      octavio augusto diaz trejo             Mexico

                   Region  Number_of_Bags Bag.Weight  \
21             san ramon           250.0      3 lbs
30                 huila           275.0          6
34                oromia           320.0      60 kg
35               muranga           320.0       1 kg
43          natou county            10.0      15 kg
...                   ...             ...        ...
1167               huila           250.0      70 kg
1182        natou county            50.0      20 kg
1183            veracruz            12.0       1 kg
1205  zaragoza itundujia            14.0       1 kg
1209            veracruz            20.0       1 kg

                      In.Country.Partner
Owner_1  \
21          Specialty Coffee Association          The Coffee Source
Inc.
30          Specialty Coffee Association                       Nora
Zeas
34    METAD Agricultural Development plc                  Hider
Abamecha
35      Kenya Coffee Traders Association                    Daniel
Magu
43          Specialty Coffee Association          Lin, Che-Hao Krude
林哲豪
...                                    ...                              .
..
1167                           Almacafé  Exportadora de Cafe Condor
S.A
1182        Specialty Coffee Association          Lin, Che-Hao Krude
林哲豪
1183                            AMECAFE    MARCO VIRGILIO RAMIREZ
TELIZ
1205                            AMECAFE       ISRAEL EDUARDO PAZ
GARCIA
1209                            AMECAFE      OCTAVIO AUGUSTO DIAZ
TREJO

               ProcessingMethod  Aroma  ...  Clean.Cup  Sweetness  \
21                 Washed / Wet   8.08  ...       10.0       10.0
30                 Washed / Wet   7.92  ...       10.0       10.0
34                Natural / Dry   8.00  ...       10.0       10.0
35                 Washed / Wet   8.08  ...       10.0       10.0
```

```
43    Semi-washed / Semi-pulped   8.08  ...       10.0       10.0
...                          ...    ...  ...        ...        ...
1167               Washed / Wet   7.25  ...       10.0       10.0
1182               Washed / Wet   7.08  ...       10.0       10.0
1183               Washed / Wet   7.00  ...       10.0       10.0
1205               Washed / Wet   7.50  ...       10.0       10.0
1209               Washed / Wet   7.25  ...       10.0       10.0

      Cupper.Points  Total.Cup.Points  Moisture  Category.One.Defects
\
21             8.33             87.17      0.11                     0

30             8.00             86.58      0.08                     0

34             8.08             86.25      0.10                     0

35             8.08             86.25      0.12                     0

43             8.42             86.08      0.12                     0

...             ...               ...       ...                   ...

1167           6.92             79.58      0.10                     0

1182           6.75             79.25      0.11                     0

1183           6.92             79.25      0.13                     0

1205           6.75             78.92      0.16                     0

1209           6.67             78.75      0.14                     0


      Quakers  Category.Two.Defects
Certification.Body  \
21        0.0                     2        Specialty Coffee
Association
30        0.0                     2        Specialty Coffee
Association
34        0.0                     3  METAD Agricultural Development
plc
35        0.0                     1    Kenya Coffee Traders
Association
43        0.0                     0        Specialty Coffee
Association
...       ...                   ...                                ..
.
1167      0.0                     4
Almacafé
1182      0.0                     0        Specialty Coffee
```

```
Association
1183      0.0                     10
AMECAFE
1205      0.0                      0
AMECAFE
1209      0.0                      0
AMECAFE

      unit_of_measurement
21                      m
30                      m
34                      m
35                      m
43                      m
...                   ...
1167                    m
1182                    m
1183                    m
1205                    m
1209                    m

[592 rows x 26 columns]
```

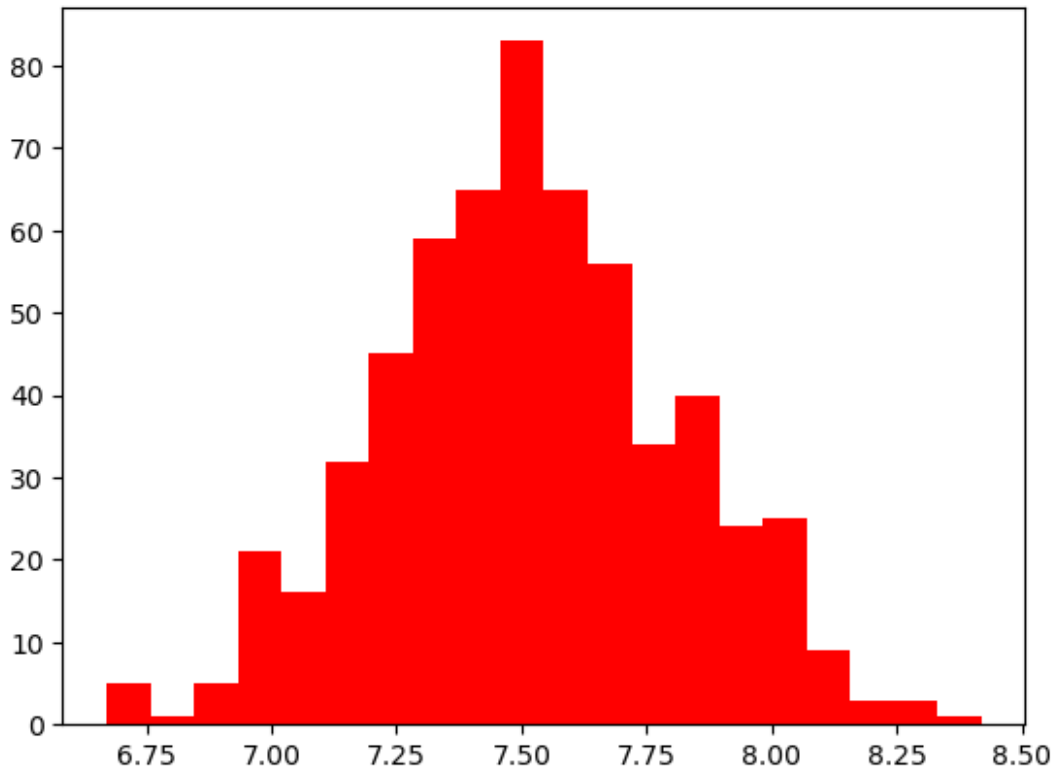
```
filter.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 592 entries, 21 to 1209
Data columns (total 26 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Species               592 non-null    object
 1   Owner                 592 non-null    object
 2   Country_of_Origin     592 non-null    object
 3   Region                592 non-null    object
 4   Number_of_Bags        592 non-null    float64
 5   Bag.Weight            592 non-null    object
 6   In.Country.Partner    592 non-null    object
 7   Owner_1               592 non-null    object
 8   ProcessingMethod      592 non-null    object
 9   Aroma                 592 non-null    float64
 10  Flavor                592 non-null    float64
 11  Aftertaste            592 non-null    float64
 12  Acidity               592 non-null    float64
 13  Body                  592 non-null    float64
 14  Balance               592 non-null    float64
 15  Uniformity            592 non-null    float64
 16  Clean.Cup             592 non-null    float64
 17  Sweetness             592 non-null    float64
 18  Cupper.Points         592 non-null    float64
 19  Total.Cup.Points      592 non-null    float64
```

```
 20  Moisture              592 non-null    float64
 21  Category.One.Defects  592 non-null    int64
 22  Quakers               592 non-null    float64
 23  Category.Two.Defects  592 non-null    int64
 24  Certification.Body    592 non-null    object
 25  unit_of_measurement   592 non-null    object
dtypes: float64(14), int64(2), object(10)
memory usage: 124.9+ KB
```

```
#Univariate analysis
import matplotlib.pyplot as plt
plt.hist(filter['Cupper.Points'],bins=20,color='red')
```

```
(array([ 5.,  1.,  5., 21., 16., 32., 45., 59., 65., 83., 65., 56.,
34.,
        40., 24., 25.,  9.,  3.,  3.,  1.]),
 array([6.67  , 6.7575, 6.845 , 6.9325, 7.02  , 7.1075, 7.195 ,
7.2825,
        7.37  , 7.4575, 7.545 , 7.6325, 7.72  , 7.8075, 7.895 ,
7.9825,
        8.07  , 8.1575, 8.245 , 8.3325, 8.42  ]),
 <BarContainer object of 20 artists>)
```

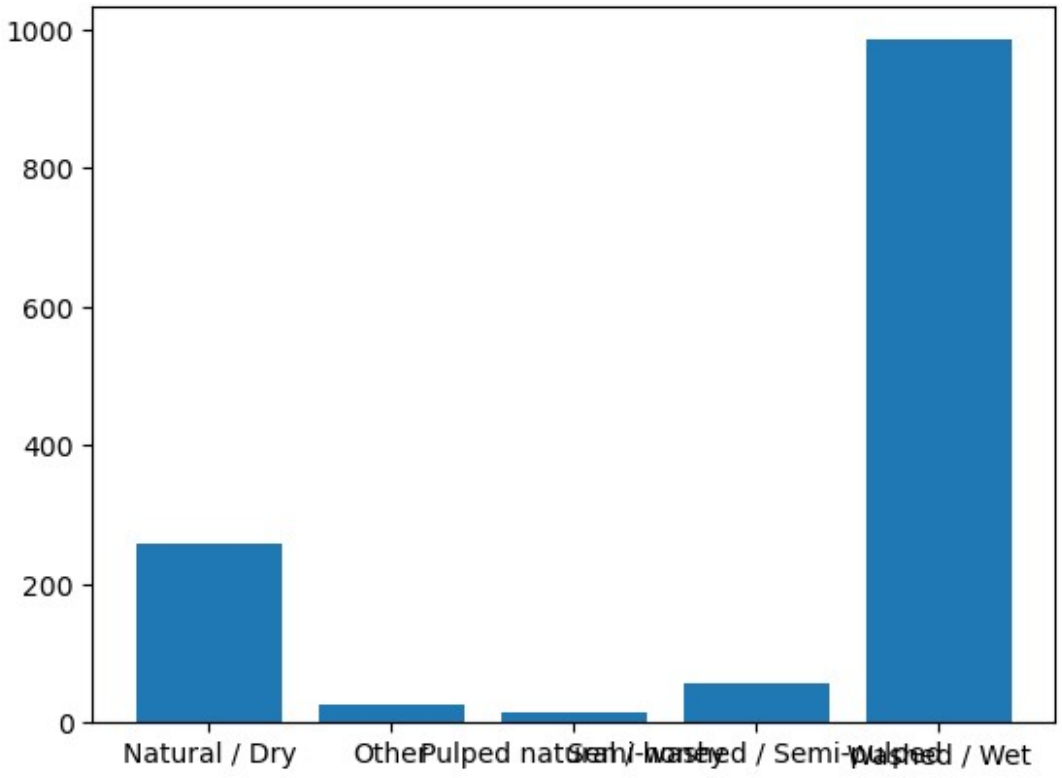


```
e=c.groupby(['ProcessingMethod']).size().reset_index(name="count").ren
ame(columns={'ProcessingMethod':'Processing_Method'})
e
```

	Processing_Method	count
0	Natural / Dry	258
1	Other	26
2	Pulped natural / honey	14
3	Semi-washed / Semi-pulped	56
4	Washed / Wet	985

```
plt.bar(e['Processing_Method'],e['count'])
```

```
<BarContainer object of 5 artists>
```



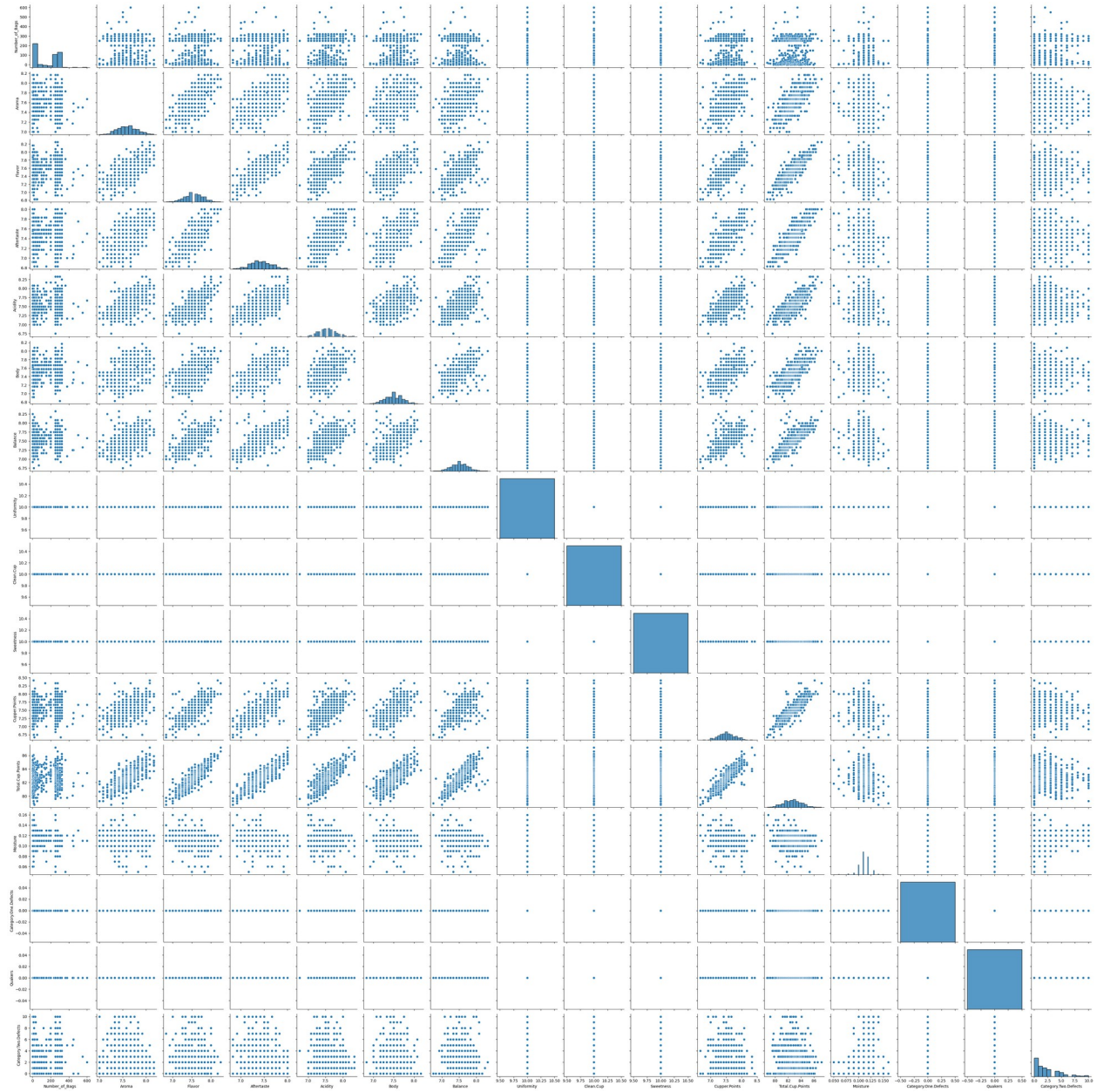

```
#Adding an extra column "count%" to show the percentage of each
processing method present
e['count%']=e['count']/sum(e['count'])*100
e
```

	Processing_Method	count	count%
0	Natural / Dry	258	19.268111
1	Other	26	1.941748
2	Pulped natural / honey	14	1.045556

```
3  Semi-washed / Semi-pulped     56   4.182226
4                Washed / Wet    985  73.562360
```

```
#Bivariate analysis
import seaborn as sns
e=sns.pairplot(filter)
e
```

```
<seaborn.axisgrid.PairGrid at 0x1e96e731750>
```



```
filter.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 592 entries, 21 to 1209
Data columns (total 26 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   Species                592 non-null    object
 1   Owner                  592 non-null    object
 2   Country_of_Origin      592 non-null    object
 3   Region                 592 non-null    object
 4   Number_of_Bags         592 non-null    float64
 5   Bag.Weight             592 non-null    object
 6   In.Country.Partner     592 non-null    object
 7   Owner_1                592 non-null    object
 8   ProcessingMethod       592 non-null    object
 9   Aroma                  592 non-null    float64
 10  Flavor                 592 non-null    float64
 11  Aftertaste             592 non-null    float64
 12  Acidity                592 non-null    float64
 13  Body                   592 non-null    float64
 14  Balance                592 non-null    float64
 15  Uniformity             592 non-null    float64
 16  Clean.Cup              592 non-null    float64
 17  Sweetness              592 non-null    float64
 18  Cupper.Points          592 non-null    float64
 19  Total.Cup.Points       592 non-null    float64
 20  Moisture               592 non-null    float64
 21  Category.One.Defects   592 non-null    int64
 22  Quakers                592 non-null    float64
 23  Category.Two.Defects   592 non-null    int64
 24  Certification.Body     592 non-null    object
 25  unit_of_measurement    592 non-null    object
dtypes: float64(14), int64(2), object(10)
memory usage: 124.9+ KB

e=filter.select_dtypes(exclude=['object'])
e

      Number_of_Bags  Aroma  Flavor  Aftertaste  Acidity  Body
Balance  \
21             250.0   8.08    8.25        8.00     8.17  8.00
8.33
30             275.0   7.92    8.25        8.00     8.33  8.00
8.08
34             320.0   8.00    8.08        7.92     8.00  8.08
8.08
35             320.0   8.08    8.00        8.00     8.25  7.92
7.92
43              10.0   8.08    8.17        7.75     8.08  7.75
7.83
...              ...    ...     ...         ...      ...   ...      ..
```

```
.
1167           250.0   7.25    7.17         7.00     6.75  7.17
7.33
1182            50.0   7.08    6.83         6.83     7.25  7.42
7.08
1183            12.0   7.00    7.00         6.92     7.17  7.17
7.08
1205            14.0   7.50    7.00         6.92     7.08  6.92
6.75
1209            20.0   7.25    6.83         6.83     7.00  7.17
7.00

      Uniformity  Clean.Cup  Sweetness  Cupper.Points
Total.Cup.Points  \
21          10.0       10.0       10.0           8.33
87.17
30          10.0       10.0       10.0           8.00
86.58
34          10.0       10.0       10.0           8.08
86.25
35          10.0       10.0       10.0           8.08
86.25
43          10.0       10.0       10.0           8.42
86.08
...          ...        ...        ...            ...                ..
.
1167        10.0       10.0       10.0           6.92
79.58
1182        10.0       10.0       10.0           6.75
79.25
1183        10.0       10.0       10.0           6.92
79.25
1205        10.0       10.0       10.0           6.75
78.92
1209        10.0       10.0       10.0           6.67
78.75

      Moisture  Category.One.Defects  Quakers  Category.Two.Defects
21        0.11                     0      0.0                     2
30        0.08                     0      0.0                     2
34        0.10                     0      0.0                     3
35        0.12                     0      0.0                     1
43        0.12                     0      0.0                     0
...        ...                   ...      ...                   ...
1167      0.10                     0      0.0                     4
1182      0.11                     0      0.0                     0
1183      0.13                     0      0.0                    10
1205      0.16                     0      0.0                     0
1209      0.14                     0      0.0                     0
```

```
[592 rows x 16 columns]
```

```
df2=e.drop(columns=['Uniformity','Clean.Cup','Sweetness','Category.One
.Defects','Category.One.Defects'])
df2
```

```
      Number_of_Bags  Aroma  Flavor  Aftertaste  Acidity  Body
Balance  \
21             250.0   8.08    8.25        8.00     8.17  8.00
8.33
30             275.0   7.92    8.25        8.00     8.33  8.00
8.08
34             320.0   8.00    8.08        7.92     8.00  8.08
8.08
35             320.0   8.08    8.00        8.00     8.25  7.92
7.92
43              10.0   8.08    8.17        7.75     8.08  7.75
7.83
...              ...    ...     ...         ...      ...   ...       ..
.
1167           250.0   7.25    7.17        7.00     6.75  7.17
7.33
1182            50.0   7.08    6.83        6.83     7.25  7.42
7.08
1183            12.0   7.00    7.00        6.92     7.17  7.17
7.08
1205            14.0   7.50    7.00        6.92     7.08  6.92
6.75
1209            20.0   7.25    6.83        6.83     7.00  7.17
7.00

      Cupper.Points  Total.Cup.Points  Moisture  Quakers
Category.Two.Defects
21             8.33             87.17      0.11      0.0
2
30             8.00             86.58      0.08      0.0
2
34             8.08             86.25      0.10      0.0
3
35             8.08             86.25      0.12      0.0
1
43             8.42             86.08      0.12      0.0
0
...             ...               ...       ...       ...
...
1167           6.92             79.58      0.10      0.0
4
1182           6.75             79.25      0.11      0.0
0
```

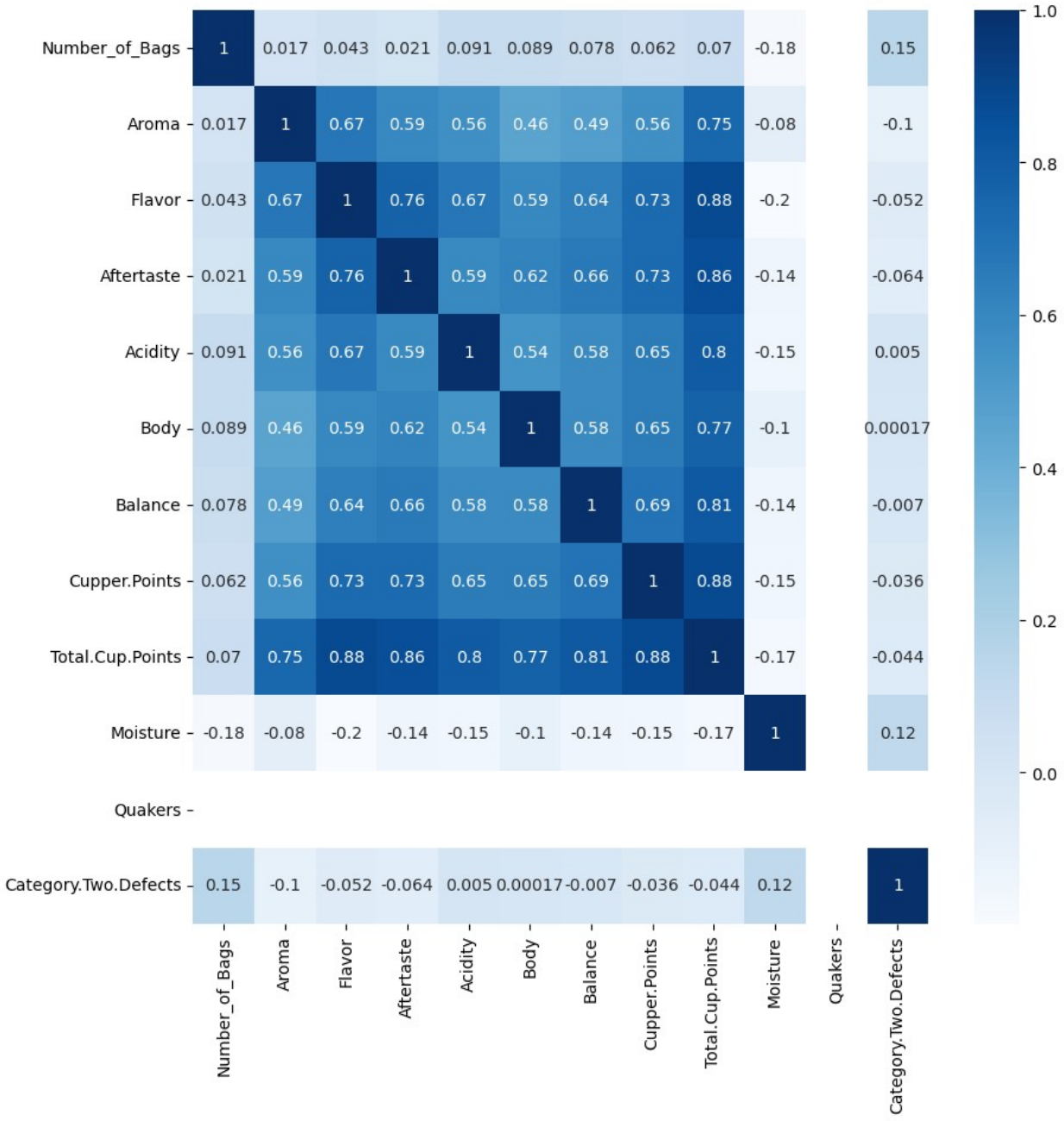
```
1183           6.92              79.25      0.13      0.0
10
1205           6.75              78.92      0.16      0.0
0
1209           6.67              78.75      0.14      0.0
0

[592 rows x 12 columns]
```

```
df2.corr()
```

```
                      Number_of_Bags     Aroma    Flavor
Aftertaste  \
Number_of_Bags              1.000000  0.016714  0.043065    0.021425

Aroma                       0.016714  1.000000  0.670133    0.593727

Flavor                      0.043065  0.670133  1.000000    0.761918

Aftertaste                  0.021425  0.593727  0.761918    1.000000

Acidity                     0.091179  0.558948  0.671595    0.585304

Body                        0.089487  0.456227  0.594085    0.618701

Balance                     0.077773  0.488713  0.635438    0.662380

Cupper.Points               0.061557  0.556487  0.730047    0.730731

Total.Cup.Points            0.069869  0.746388  0.880839    0.861924

Moisture                   -0.180161 -0.080026 -0.198329   -0.143079

Quakers                          NaN       NaN       NaN         NaN

Category.Two.Defects        0.152571 -0.101692 -0.051657   -0.064072


                       Acidity      Body   Balance  Cupper.Points  \
Number_of_Bags        0.091179  0.089487  0.077773       0.061557
Aroma                 0.558948  0.456227  0.488713       0.556487
Flavor                0.671595  0.594085  0.635438       0.730047
Aftertaste            0.585304  0.618701  0.662380       0.730731
Acidity               1.000000  0.537732  0.582737       0.649262
Body                  0.537732  1.000000  0.580287       0.649705
Balance               0.582737  0.580287  1.000000       0.688170
Cupper.Points         0.649262  0.649705  0.688170       1.000000
Total.Cup.Points      0.799773  0.766142  0.809651       0.878995
Moisture             -0.154790 -0.102323 -0.142152      -0.154504
Quakers                    NaN       NaN       NaN            NaN
Category.Two.Defects  0.004964  0.000167 -0.006968      -0.035668
```

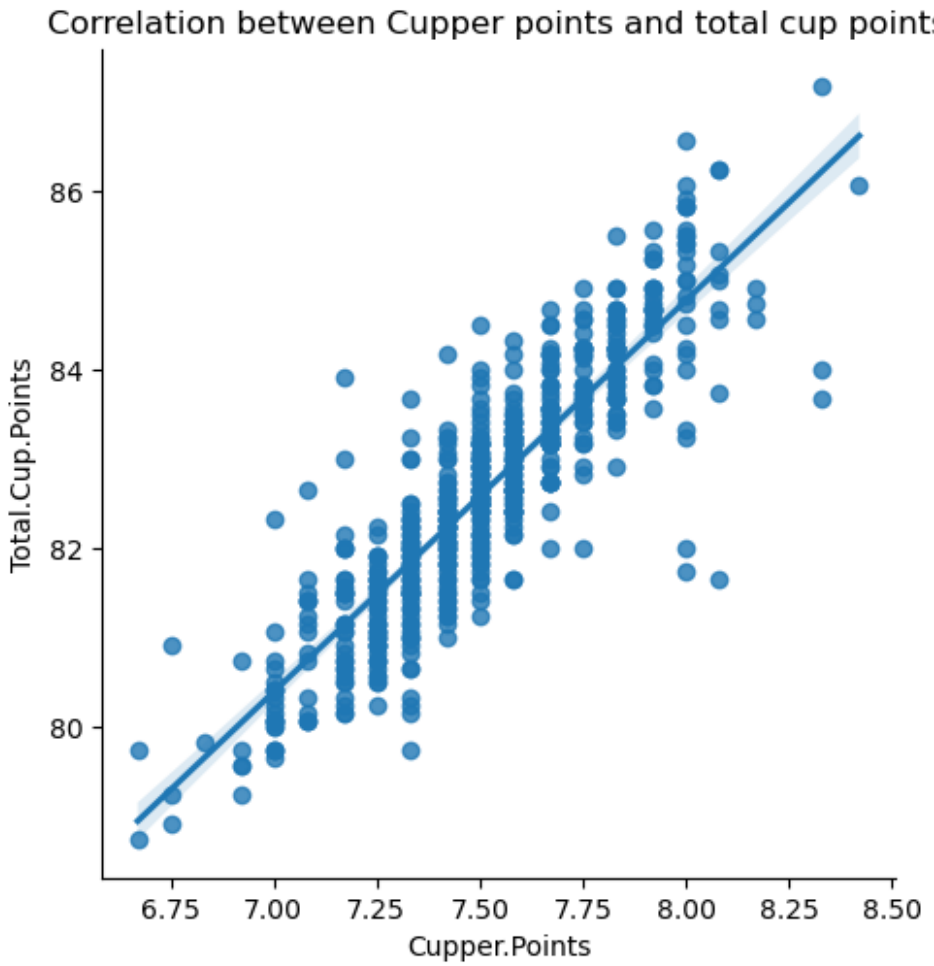
```
                      Total.Cup.Points  Moisture  Quakers  \
Number_of_Bags                0.069869 -0.180161      NaN
Aroma                         0.746388 -0.080026      NaN
Flavor                        0.880839 -0.198329      NaN
Aftertaste                    0.861924 -0.143079      NaN
Acidity                       0.799773 -0.154790      NaN
Body                          0.766142 -0.102323      NaN
Balance                       0.809651 -0.142152      NaN
Cupper.Points                 0.878995 -0.154504      NaN
Total.Cup.Points              1.000000 -0.170731      NaN
Moisture                     -0.170731  1.000000      NaN
Quakers                            NaN       NaN      NaN
Category.Two.Defects         -0.043774  0.124455      NaN

                      Category.Two.Defects
Number_of_Bags                    0.152571
Aroma                            -0.101692
Flavor                           -0.051657
Aftertaste                       -0.064072
Acidity                           0.004964
Body                              0.000167
Balance                          -0.006968
Cupper.Points                    -0.035668
Total.Cup.Points                 -0.043774
Moisture                          0.124455
Quakers                                NaN
Category.Two.Defects              1.000000
```

```
#Plotting a heatmap for better visualization of the correlation
matrix, to get an intuitive representation of relationships between
variables
plt.figure(figsize=(10,10))
sns.heatmap(df2.corr() ,annot=True,cmap='Blues')
plt.show()
```



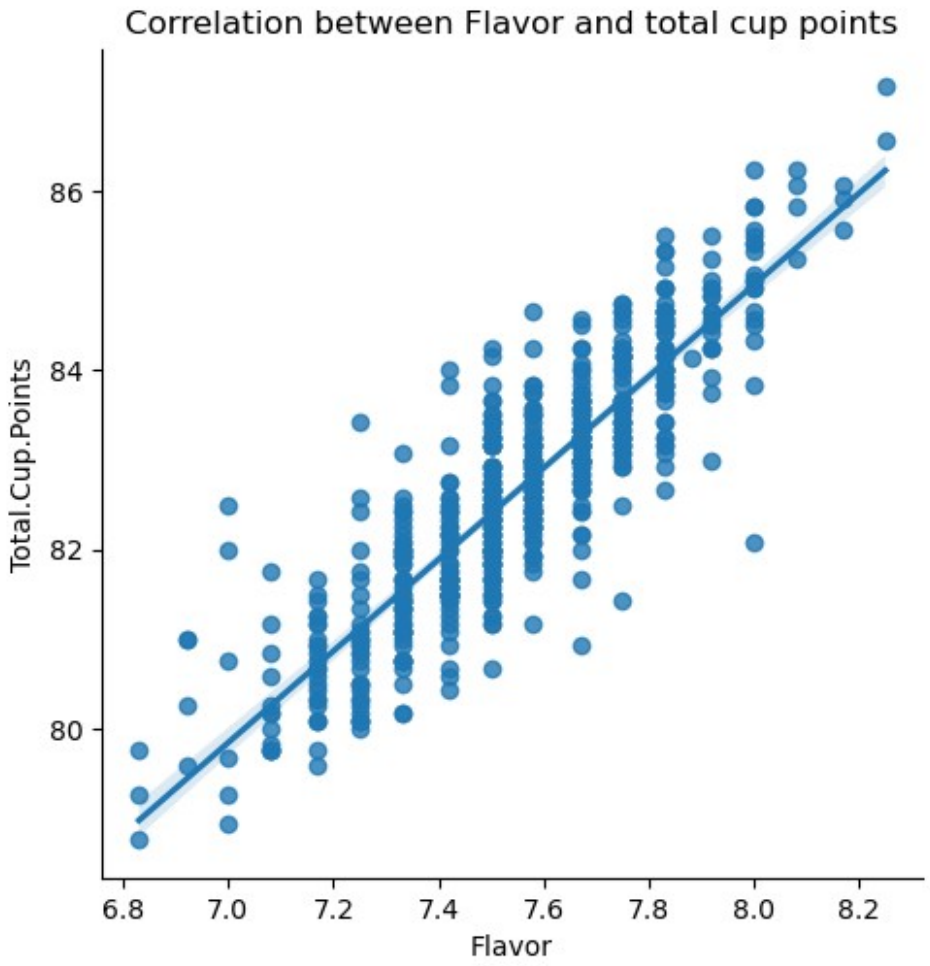

```
sns.lmplot(x='Cupper.Points',y='Total.Cup.Points',data=df2)
plt.title("Correlation between Cupper points and total cup points")
```

```
Text(0.5, 1.0, 'Correlation between Cupper points and total cup
points')
```




```
sns.lmplot(x='Flavor',y='Total.Cup.Points',data=df2)
plt.title("Correlation between Flavor and total cup points")
```

```
Text(0.5, 1.0, 'Correlation between Flavor and total cup points')
```




```
#Assign the independent variables to x and the dependent variable
(total cup points) to y for performing ANOVA test
x=filter[['Flavor','Acidity','Balance','Cupper.Points']]
y=filter['Total.Cup.Points']

x
```

	Flavor	Acidity	Balance	Cupper.Points
21	8.25	8.17	8.33	8.33
30	8.25	8.33	8.08	8.00
34	8.08	8.00	8.08	8.08
35	8.00	8.25	7.92	8.08
43	8.17	8.08	7.83	8.42
...	...	...	...	...
1167	7.17	6.75	7.33	6.92
1182	6.83	7.25	7.08	6.75
1183	7.00	7.17	7.08	6.92
1205	7.00	7.08	6.75	6.75
1209	6.83	7.00	7.00	6.67

```
[592 rows x 4 columns]
```

```
y
```

```
21      87.17
30      86.58
34      86.25
35      86.25
43      86.08
        ...
1167    79.58
1182    79.25
1183    79.25
1205    78.92
1209    78.75
Name: Total.Cup.Points, Length: 592, dtype: float64
```

```
from sklearn.feature_selection import f_classif
f=f_classif(x,y)
f
```

```
(array([25.53697691, 14.16599345, 14.58635089, 25.4364055 ]),
 array([1.03140812e-134, 2.76042872e-088, 2.19215384e-090,
2.24140767e-134]))
```

```
#So we can see that Quality of the coffee is more dependable on the
Flavour and Cupper.Points
```